

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 06:16:44 ; Search time 12.13 Seconds  
(without alignments)  
111.269 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 67  
Sequence: 1 GGFGLGRGKCPNSEIFSR.....CRGLRNKKKVCPRSKCG 67

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 193259 seqs, 2014635 residues

Word size: 0

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database:

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	11.9	738	3	US-08-864-038A-3
2	7	10.4	47	2	US-08-637-759B-400
3	7	10.4	47	3	US-08-871-355A-400
4	6	9.0	130	2	US-08-888-497-43
5	6	9.0	130	5	PCT-US94-07926-43
6	6	9.0	158	2	US-08-888-497-22
7	6	9.0	158	5	PCT-US94-07926-22
8	6	9.0	253	1	US-08-399-696-4
9	6	9.0	253	1	US-08-399-696-118
10	6	9.0	400	1	US-07-730-953-2
11	6	9.0	488	2	US-08-928-692-10
12	6	9.0	509	1	US-08-030-096-2
13	6	9.0	568	4	US-09-160-065-2
14	6	9.0	590	4	US-09-413-814-89
15	6	9.0	591	4	US-09-413-814-76
16	6	9.0	2182	2	US-08-487-826B-16
17	6	9.0	3075	2	US-08-460-309-5
18	6	9.0	3075	2	US-08-125-077-5
19	5	7.5	5	3	US-09-026-633-2
20	5	7.5	5	3	US-08-929-329-12
21	5	7.5	8	4	US-09-128-450-4
22	5	7.5	9	1	US-08-423-069-11
23	5	7.5	9	2	US-08-317-844B-11
24	5	7.5	9	4	US-08-963-168C-41
25	5	7.5	9	4	US-08-963-168C-42
26	5	7.5	9	4	US-09-353-976-3
27	5	7.5	9	4	US-09-353-976-5

28	5	7.5	9	4	US-09-353-976-6	Sequence 6, Appl1
29	5	7.5	9	4	US-09-258-754-67	Sequence 67, Appl1
30	5	7.5	9	4	US-09-042-107-67	Sequence 67, Appl1
31	5	7.5	10	2	US-08-724-548-40	Sequence 40, Appl1
32	5	7.5	10	2	US-08-724-548-41	Sequence 41, Appl1
33	5	7.5	10	2	US-08-724-548-42	Sequence 42, Appl1
34	5	7.5	10	2	US-08-724-548-43	Sequence 43, Appl1
35	5	7.5	10	2	US-08-724-548-44	Sequence 44, Appl1
36	5	7.5	10	2	US-08-724-548-45	Sequence 45, Appl1
37	5	7.5	10	3	US-07-978-674B-40	Sequence 40, Appl1
38	5	7.5	10	3	US-07-978-674B-41	Sequence 41, Appl1
39	5	7.5	10	3	US-07-978-674B-42	Sequence 42, Appl1
40	5	7.5	10	3	US-07-978-674B-43	Sequence 43, Appl1
41	5	7.5	10	3	US-07-978-674B-44	Sequence 44, Appl1
42	5	7.5	10	3	US-07-978-674B-45	Sequence 45, Appl1
43	5	7.5	10	4	US-08-963-168C-10	Sequence 10, Appl1
44	5	7.5	10	4	US-08-963-168C-17	Sequence 17, Appl1
45	5	7.5	10	4	US-09-128-450-6	Sequence 6, Appl1
46	5	7.5	13	1	US-08-425-069-39	Sequence 39, Appl1
47	5	7.5	13	2	US-08-317-844B-39	Sequence 39, Appl1
48	5	7.5	14	2	US-08-685-589A-166	Sequence 166, App
49	5	7.5	15	1	US-08-425-069-13	Sequence 13, Appl1
50	5	7.5	15	2	US-08-317-844B-13	Sequence 13, Appl1

## ALIGNMENTS

RESULT 1  
US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano  
; STREET: Isshiniden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pinctada fucata

CELL TYPE: mantle epithelial cell  
FEATURE:  
NAME/KEY: peptide  
LOCATION: from 1 to 738  
IDENTIFICATION METHOD: E (by experiment)  
US-08-864-038A-3

Query Match 11.9%; Score 8; DB 3; Length 738;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGLGC 8  
|||||  
Db 493 GGGGGLGC 500

RESULT 2  
US-08-637-759B-400  
Sequence 400, Application US/08637759B  
Patent No. 5876931  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/637,759B  
FILING DATE: 03-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 400:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-637-759B-400

Query Match 10.4%; Score 7; DB 2; Length 47;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SRCDGRC 25  
|||||  
Db 37 SRCDGRC 43

RESULT 3  
US-08-871-355A-400  
Sequence 400, Application US/08871355A  
Patent No. 6015669  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,355A  
FILING DATE: 09-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101 CON  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 400:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-871-355A-400

Query Match 10.4%; Score 7; DB 3; Length 47;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SRCDGRC 25  
|||||  
Db 37 SRCDGRC 43

RESULT 4  
US-08-888-497-43  
Sequence 43, Application US/08888497  
Patent No. 5972677  
GENERAL INFORMATION:  
APPLICANT: Tisfield, Jay A.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
SEQUENCES AND LOW MOLECULAR WEIGHT AMINO ACID SEQUENCES  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
ADDRESSEE: Russell PA  
STREET: 200 East Broadway Boulevard  
CITY: Fort Lauderdale

\* STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/888,497  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,405  
FILING DATE:  
APPLICATION NUMBER: US 08/097,354  
FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-764-4996  
TELEFAX: 305-527-2498  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 130 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-888-497-43

Query Match  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10  
|||||  
Db 29 GLGGRG 34

RESULT 5  
PCT-US94-07926-43  
Sequence 43, Application PC/TUS94/07926  
GENERAL INFORMATION:  
APPLICANT: Tischfield, Jay A.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
STREET: 200 East Broward Boulevard  
CITY: Fort Lauderdale  
STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07926  
FILING DATE: 15-JUL-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/097,354

FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-764-4996  
TELEFAX: 305-527-2498  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 130 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-07926-43

Query Match  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10  
|||||  
Db 29 GLGGRG 34

RESULT 6  
US-08-888-497-22  
Sequence 22, Application US/08888497  
Patent No. 5972677  
GENERAL INFORMATION:  
APPLICANT: Tischfield, Jay A.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
STREET: 200 East Broward Boulevard  
CITY: Fort Lauderdale  
STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/888,497  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,405  
FILING DATE:  
APPLICATION NUMBER: US 08/097,354  
FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-764-4996  
TELEFAX: 305-527-2498  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 158 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-888-497-22

Query Match 9.0%; Score 6; DB 2; Length 158;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10  
|||||  
DB 57 GLGGRG 62

RESULT 7  
PCT-US94-07926-22  
Sequence 22, Application PC/TUS9407926  
GENERAL INFORMATION:  
APPLICANT: Tischfield, Jay A.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
ADDRESS: Russell PA  
STREET: 200 East Broward Boulevard  
CITY: Fort Lauderdale  
STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07926  
FILING DATE: 15-JUL-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/097,354  
FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-527-2498  
TELEFAX: 305-764-4996  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 158 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-07926-22

Query Match 9.0%; Score 6; DB 5; Length 158;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10  
|||||  
DB 57 GLGGRG 62

RESULT 8  
US-08-399-696-4  
Sequence 4, Application US/08399696  
Patent No. 5756669

GENERAL INFORMATION:  
APPLICANT: p53-BINDING POLYPEPTIDES AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING SAME  
NUMBER OF SEQUENCES: 126  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,696  
FILING DATE: 02-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/156,671  
FILING DATE: 22-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15522-000710  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids

US-08-399-696-4  
MOLECULE TYPE: protein

QY 4 GGLGR 9  
|||||  
DB 199 GGLGR 204

Query Match 9.0%; Score 6; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-399-696-118  
Sequence 118, Application US/08399696  
Patent No. 5756669  
GENERAL INFORMATION:  
APPLICANT: p53-BINDING POLYPEPTIDES AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING SAME  
NUMBER OF SEQUENCES: 126  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,696  
FILING DATE: 02-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/156,671  
FILING DATE: 22-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15522-000710  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids

TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-399-696-118

Query Match 9.0%; Score 6; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGR 9  
|||||  
DB 199 GGLGR 204

RESULT 10  
US-07-730-953-2  
Sequence 2, Application US/07730953  
Patent No. 5286614

GENERAL INFORMATION:  
APPLICANT: BODENMULLER, Heinz  
APPLICANT: DESSAUER, Andreas  
TITLE OF INVENTION: METHOD FOR THE DETECTION OF MALIGNANT  
TITLE OF INVENTION: DISEASES  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/730,953  
FILING DATE: 19910723  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 39 42 999.7  
FILING DATE: 21-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Kites, Monica C.  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P564-1119  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)638-5000  
TELEFAX: (202)638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-730-953-2

Query Match 9.0%; Score 6; DB 1; Length 400;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FGGIGG 8  
|||||  
DB 15 FGGIGG 20

RESULT 11  
US-08-928-692-10  
Sequence 10, Application US/08928692

Patent No. 5958727  
GENERAL INFORMATION:  
APPLICANT: Brody, Howard  
APPLICANT: Yaver, Deborah S.  
APPLICANT: Lamsa, Michael  
APPLICANT: Hansen, Kim  
TITLE OF INVENTION: Methods for Modifying the Production of  
TITLE OF INVENTION: a Polypeptide  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59587270 No. 5958727/disk of No. 5958727th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,692  
FILING DATE: 12-SEPT-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4944.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5958727e  
US-08-928-692-10

Query Match 9.0%; Score 6; DB 2; Length 488;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 LGGRGK 11  
|||||  
DB 441 LGGRGK 446

RESULT 12  
US-08-030-096-2  
Sequence 2, Application US/08030096  
Patent No. 5426041

GENERAL INFORMATION:  
APPLICANT: Fabijanski, Steven F.  
APPLICANT: Arnison, Paul G.  
TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/030,096  
 FILING DATE: 22-MAR-1993  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/556,917  
 FILING DATE: 20-JUL-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA91/00235  
 FILING DATE: 22-JUL-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 33229/164/PIHI  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 509 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-030-096-2

Query Match 9.0%; Score 6; DB 1; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGL 6  
 |||||  
 DB 94 GGGGGL 99

RESULT 13  
 US-09-160-065-2  
 ; Sequence 2, Application US/09160065  
 ; Patent No. 6146641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lee, Lucy  
 ; APPLICANT: Fady, Aly  
 ; APPLICANT: Hunt, Henry  
 ; TITLE OF INVENTION: Avian Leukosis Virus Subgroup J Envelope Gene Product  
 ; FILE REFERENCE: Docket No. 6146641 0155.98 - Lee, Lucy  
 ; CURRENT APPLICATION NUMBER: US/09/160,065  
 ; CURRENT FILING DATE: 1998-09-24  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: Patentln Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 568  
 ; TYPE: PRT  
 ; ORGANISM: Avian Leukosis virus  
 US-09-160-065-2

Query Match 9.0%; Score 6; DB 4; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FGGGIG 8  
 |||||  
 DB 508 FGGGIG 513

RESULT 14  
 US-09-413-814-89  
 ; Sequence 89, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan  
 APPLICANT: Bloeker, Helmut  
 APPLICANT: Brandt, Petra  
 APPLICANT: Cino, Paul M  
 APPLICANT: Dougherty, Brian A  
 APPLICANT: Goldberg, Steven L  
 APPLICANT: Hofle, Gerhard  
 APPLICANT: Mueller, Joachim  
 APPLICANT: Reichenbach, Hans  
 TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
 FILE REFERENCE: PCT/US 99/23535  
 CURRENT APPLICATION NUMBER: US/09/413,814  
 CURRENT FILING DATE: 1999-10-07  
 EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 EARLIER FILING DATE: 1998-10-09  
 NUMBER OF SEQ ID NOS: 107  
 SOFTWARE: Patentln Ver. 2.1  
 SEQ ID NO 89  
 LENGTH: 590  
 TYPE: PRT  
 ORGANISM: Sorangium cellulosum  
 US-09-413-814-89

Query Match 9.0%; Score 6; DB 4; Length 590;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 VCVPRS 64  
 |||||  
 DB 145 VCVPRS 150

RESULT 15  
 US-09-413-814-76  
 ; Sequence 76, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Beyer, Stefan  
 ; APPLICANT: Bloeker, Helmut  
 ; APPLICANT: Brandt, Petra  
 ; APPLICANT: Cino, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hofle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: US/09/413,814  
 ; CURRENT FILING DATE: 1999-10-07  
 ; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 ; EARLIER FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: Patentln Ver. 2.1  
 ; SEQ ID NO 76  
 ; LENGTH: 591  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-413-814-76

Query Match 9.0%; Score 6; DB 4; Length 591;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 VCVPRS 64  
 |||||  
 DB 146 VCVPRS 151

RESULT 16  
US-08-487-826B-16  
; Sequence 16, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chittis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellem, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2182 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; US-08-487-826B-16

Query Match 9.0%; Score 6; DB 2; Length 2182;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 NVVXP 36  
| | | | |  
DB 1684 NVVXP 1689

RESULT 17  
US-08-460-309-5  
; Sequence 5, Application US/08460309  
; Patent No. 5837496  
; GENERAL INFORMATION:  
; APPLICANT: Engvall, Eva  
; APPLICANT: Leivo, Ilmo  
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin  
; TITLE OF INVENTION: Fragments and Uses Thereof  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,309  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/125,077  
; FILING DATE: 22-SEP-1993  
; APPLICATION NUMBER: US PCT/US 94/10730  
; FILING DATE: 21-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/472,319  
; FILING DATE: 30-JAN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/919,951  
; FILING DATE: 27-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 9721  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3075 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-460-309-5

Query Match 9.0%; Score 6; DB 2; Length 3075;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 CVCRLG 51  
| | | | |  
DB 1528 CVCRLG 1533

RESULT 18  
US-08-125-077-5  
; Sequence 5, Application US/08125077  
; Patent No. 5872231  
; Patent No. 5872231 5840863  
; GENERAL INFORMATION:  
; APPLICANT: Engvall, Eva  
; APPLICANT: Leivo, Ilmo  
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin  
; TITLE OF INVENTION: Fragments and Uses Thereof  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/125,077  
FILING DATE: 22-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US 94/10730  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/472,319  
FILING DATE: 30-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/919,951  
FILING DATE: 27-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 9721  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9601  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3075 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-125-077-5

Query Match  
Best Local Similarity 100.0%; Score 6; DB 2; Length 3075;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 CVCRLG 51  
DB 1528 CVCRLG 1533

RESULT 19  
US-09-026-633-2  
Sequence 2, Application US/09026633  
Patent No. 6025328  
GENERAL INFORMATION:  
APPLICANT: McMorris, Trevor C.  
APPLICANT: Kelner, Michael J.  
TITLE OF INVENTION: Antitumor agents  
FILE REFERENCE: 103.008051  
CURRENT APPLICATION NUMBER: US/09/026,633  
CURRENT FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 5  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Amino acid sequence  
US-09-026-633-2

Query Match  
Best Local Similarity 7.5%; Score 5; DB 3; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CDGRC 25  
DB 1 CDGRC 5

RESULT 20  
US-08-929-329-12  
Sequence 12, Application US/08929329  
Patent No. 6120770  
GENERAL INFORMATION:  
APPLICANT: Adams, John H

APPLICANT: Dalton, John P  
APPLICANT: Kappe, Stefan  
TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing  
TITLE OF INVENTION: Vaccine Compositions  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Barnes & Thornburg  
STREET: 11 S Meridian  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/929,329  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Breen, John P  
REGISTRATION NUMBER: 38,833  
REFERENCE/DOCKET NUMBER: 835910-28685  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 231-7745  
TELEFAX: (317) 231-7743  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-08-929-329-12

Query Match  
Best Local Similarity 7.5%; Score 5; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GIGGR 9  
DB 1 GIGGR 5

RESULT 21  
US-09-128-450-4  
Sequence 4, Application US/09128450  
Patent No. 6211149  
GENERAL INFORMATION:  
APPLICANT: Chesebro, Bruce W  
APPLICANT: Caughey, Byron W  
APPLICANT: Chabry, Joelle  
APPLICANT: Priola, Susette  
TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion  
TITLE OF INVENTION: Protein  
FILE REFERENCE: 50121  
CURRENT APPLICATION NUMBER: US/09/128,450  
CURRENT FILING DATE: 1998-08-03  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Hamster sp.  
US-09-128-450-4



Query Match 7.5%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLG 8  
| | | | |  
DB 3 GGLG 7

## RESULT 22

US-08-425-069-11

Sequence 11, Application US/08425069

Patent No. 572810

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V.

APPLICANT: Xu, Ming

APPLICANT: Hinman, Michael B.

TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK

TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL

TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch &amp; Birch

STREET: 301 No. 572810th Washington Street

CITY: Falls Church

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22046

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/425,069

FILING DATE: 19-APR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-106P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX:

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Nephila clavipes

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..9

OTHER INFORMATION: /label=1st segment

OTHER INFORMATION: /note="first segment of spider silk protein"

OTHER INFORMATION: repeats."

US-08-425-069-11

Query Match 7.5%; Score 5; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLG 8  
| | | | |  
DB 4 GGLG 8

## RESULT 23

US-08-317-844B-11  
Sequence 11, Application US/08317844B  
Patent No. 5969894

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V.

APPLICANT: Xu, Ming

APPLICANT: Hinman, Michael B.

TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK

TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL

TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF

NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch &amp; Birch

STREET: 301 No. 5969894th Washington Street

CITY: Falls Church

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22046

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/317,844B

FILING DATE: 04-OCT-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 241-1300

TELEFAX: (703) 241-2848

TELEX: 248345

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Nephila clavipes

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..9

OTHER INFORMATION: /label=1st segment

OTHER INFORMATION: /note="first segment of spider silk protein"

OTHER INFORMATION: repeats."

US-08-317-844B-11

Query Match 7.5%; Score 5; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLG 8  
| | | | |  
DB 4 GGLG 8

## RESULT 24

US-08-963-168C-41

Sequence 41, Application US/08963168C

Patent No. 6127166

GENERAL INFORMATION:

APPLICANT: Bayley, Hagan

APPLICANT: Cao, Qunjiang

APPLICANT: Wang, Yunjuan

TITLE OF INVENTION: MOULTUSCAN LIGAMENT POLYPEPTIDES

TITLE OF INVENTION: AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/963,168C  
 FILING DATE: 03-NOV-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fasse, Peter J.  
 REGISTRATION NUMBER: 32,983  
 REFERENCE/DOCKET NUMBER: 07917/059001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 41:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 LOCATION: 9...9  
 OTHER INFORMATION: Xaa = any amino acid  
 US-08-963-168C-41

Query Match 7.5%; Score 5; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGG 5  
 |||||  
 DB 1 GGFGG 5

RESULT 25  
 US-08-963-168C-42  
 Sequence 42, Application US/08963168C  
 Patent No. 6127166  
 GENERAL INFORMATION:  
 APPLICANT: Bayley, Hagan  
 APPLICANT: Cao, Quiding  
 APPLICANT: Wang, Yunjuan  
 TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES  
 TITLE OF INVENTION: AND GENES ENCODING THEM  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/963,168C  
 FILING DATE: 03-NOV-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fasse, Peter J.  
 REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 07917/059001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 42:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 LOCATION: 3...3  
 OTHER INFORMATION: Xaa = any amino acid  
 US-08-963-168C-42

Query Match 7.5%; Score 5; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGG 5  
 |||||  
 DB 4 GGFGG 8

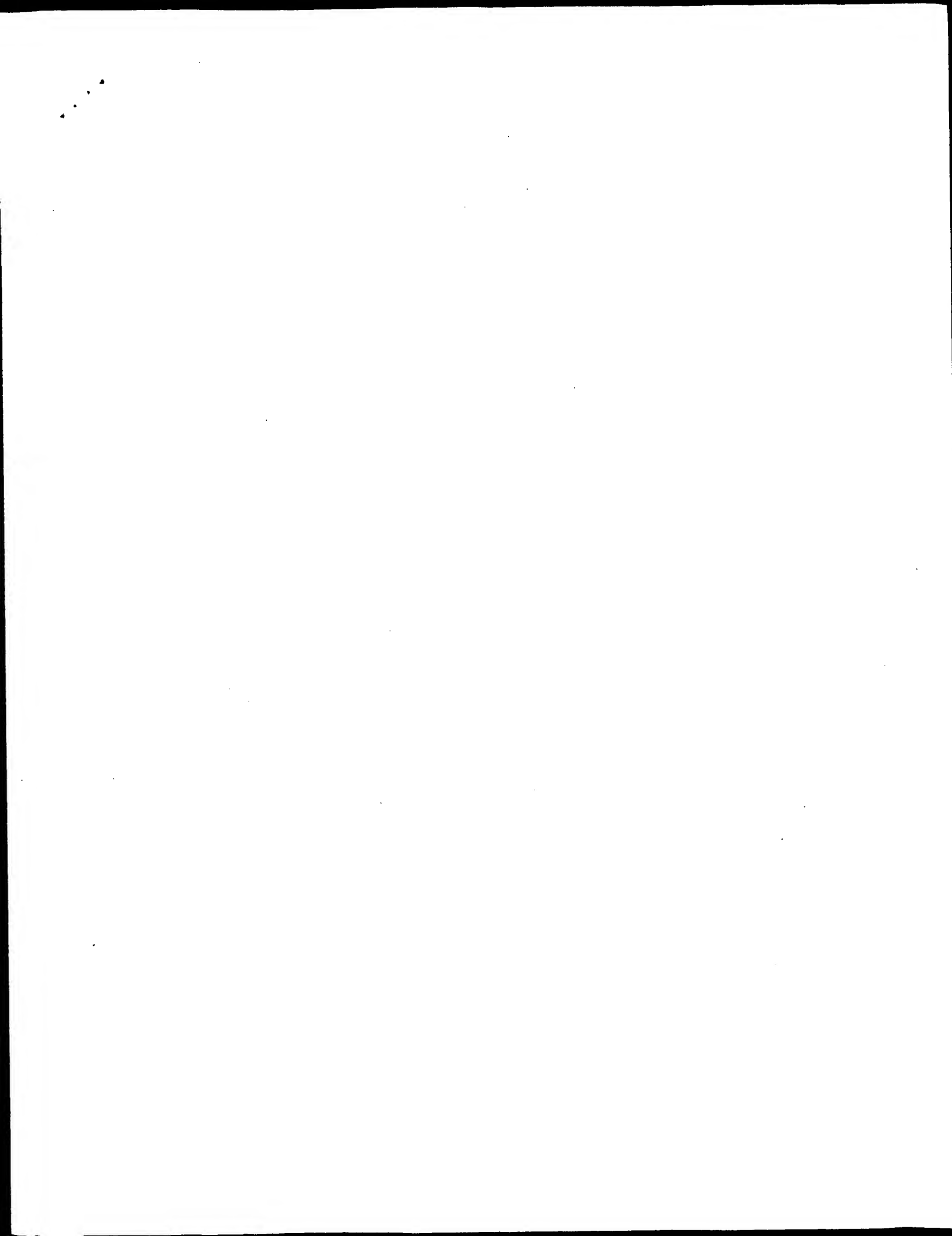
Search completed: July 12, 2001, 06:17:06  
 Job Time: 22 sec

Thu Jul 12 06:30:34 2001

us-09-506-978-1.oligo.ra1

---

Page 11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 06:17:09 ; Search time 20.97 Seconds

(without alignments)  
422.720 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 67

Sequence: 1 GFGGGLGGRGKCPSEIIFSR.....CRIGYLNKKKVCVPSKCG 67

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :  
1: SP. Archaea: \*  
2: SP. Bacteria: \*  
3: SP. Fungi: \*  
4: SP. Human: \*  
5: SP. Invertebrate: \*  
6: SP. Mammal: \*  
7: SP. MHC: \*  
8: SP. Organelle: \*  
9: SP. Phage: \*  
10: SP. Plant: \*  
11: SP. Rodent: \*  
12: SP. Unclassified: \*  
13: SP. Vertebrate: \*  
14: SP. Virus: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11.9	8	113	5 Q9V5U5	Q9V5U5 drosophila
2	11.9	8	151	10 Q22600	Q22600 onobrychis
3	11.9	8	738	5 Q02402	Q02402 pinctada fu
4	11.9	8	904	5 Q76271	Q76271 mytilus edu
5	11.9	8	922	5 Q44367	Q44367 mytilus edu
6	11.9	8	1329	2 Q06810	Q06810 mycobacteri
7	10.4	7	100	10 Q9SD66	Q9SD66 arabidopsis
8	10.4	7	104	107 Q9LEP7	Q9LEP7 brassica na
9	10.4	7	284	5 Q21073	Q21073 caenorhabdi
10	10.4	7	313	5 Q9NL74	Q9NL74 leishmania
11	10.4	7	314	5 Q76140	Q76140 trypanosoma
12	10.4	7	438	2 Q9RY66	Q9RY66 deinococcus
13	10.4	7	450	2 Q83238	Q83238 treponema p
14	10.4	7	463	2 Q10787	Q10787 mycobacteri
15	10.4	7	465	2 Q50943	Q50943 borrelia bu
16	10.4	7	556	2 Q918K4	Q918K4 mycobacteri
17	10.4	7	620	10 Q9FUI4	Q9FUI4 oryza sativ
18	10.4	7	652	2 Q68071	Q68071 rhodospirillum rubrum
19	10.4	7	682	5 Q22537	Q22537 caenorhabdi

## ALIGNMENTS

20	7	10.4	686	4 Q9H6R3	Q9H6R3 homo sapien
21	7	10.4	753	2 Q87472	Q87472 treponema p
22	7	10.4	756	2 Q83337	Q83337 treponema p
23	7	10.4	758	2 Q07894	Q07894 treponema p
24	7	10.4	758	2 Q9KHF2	Q9KHF2 treponema p
25	7	10.4	758	2 Q9KHF1	Q9KHF1 treponema p
26	7	10.4	758	2 Q9KHF0	Q9KHF0 treponema p
27	7	10.4	762	2 Q83335	Q83335 treponema p
28	7	10.4	856	1 Q58565	Q58565 pyrococcus
29	7	10.4	944	5 Q16936	Q16936 caenorhabdi
30	7	10.4	1804	10 Q9ZYV0	Q9ZYV0 arabidopsis
31	7	10.4	2160	2 Q9ZB61	Q9ZB61 proteus mir
32	7	10.4	2	2 Q07779	Q07779 mycobacteri
33	7	10.4	89	2 Q9K5X5	Q9K5X5 bacillus ha
34	7	10.4	103	5 Q9VK53	Q9VK53 drosophila
35	7	10.4	105	10 Q9LOE1	Q9LOE1 arabidopsis
36	7	10.4	113	2 Q9R645	Q9R645 streptomyces
37	7	10.4	121	10 Q9LIT1	Q9LIT1 oryza sativ
38	7	10.4	121	11 Q55213	Q55213 rattus norv
39	7	10.4	125	10 Q9SV49	Q9SV49 arabidopsis
40	7	10.4	136	10 Q9S7X8	Q9S7X8 arabidopsis
41	7	10.4	139	5 Q9N473	Q9N473 caenorhabdi
42	7	10.4	141	4 Q9H506	Q9H506 homo sapien
43	7	10.4	145	4 Q9UNK4	Q9UNK4 homo sapien
44	7	10.4	145	4 Q9UK01	Q9UK01 homo sapien
45	7	10.4	148	1 Q9YFK7	Q9YFK7 aeropyrum p
46	7	10.4	150	10 Q9S999	Q9S999 trititum ae
47	7	10.4	154	11 Q9QX68	Q9QX68 mesocricetu
48	7	10.4	165	2 Q9X9A3	Q9X9A3 streptococc
49	7	10.4	168	10 P82403	P82403 spinaclia ol
50	7	10.4	168	11 Q9QZT4	Q9QZT4 mus musculu

RESULT 1					
Q9V5U5	PRELIMINARY;	PRT;	113 AA.		
AC Q9V5U5;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)					
DE CG13217 PROTEIN.					
OS CG13217.					
OS Drosophila melanogaster (Fruit fly).					
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;					
OC Ephydroidea; Drosophilidae; Drosophila.					
OX NCBI_TaxID=7227;					
RM [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=BERKELEY.					
RX MEDLINE=20196006; PubMed=10731132;					
RA Adams M.D., Celinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,					
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,					
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,					
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,					
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,					
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,					
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,					
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,					
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,					
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,					
RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,					
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,					
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,					
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,					
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,					
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,					
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,					
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,					
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,					

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacleb J.M.,  
 RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Stimpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL EMBL: AE003826; AAF58702.1; -;  
 DR Science 287:2185-2195(2000).  
 DR Flybase: FBgn0033590; CG13217.  
 SQ SEQUENCE 113 AA; 11288 MW; 786171CE3C6AB0A3 CRC64;

Query Match 11.9%; Score 8; DB 5; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 0.52;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 66 GGFGLG 73

RESULT 2  
 ID 022600 PRELIMINARY; PRT; 151 AA.  
 AC 022600;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE GLYCINE-RICH PROTEIN.  
 OS Onchocerca vicifolia (Common salinfish).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Onchocerca.  
 OX NCBI\_TaxID=3882;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEAF;  
 RA Joseph R.G.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF027686; AAB82000.1; -;  
 DR Mengel: 25262; Oprovi:343;25262.  
 SQ SEQUENCE 151 AA; 14641 MW; 38391AC9C51087E4 CRC64;

Query Match 11.9%; Score 8; DB 10; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 76 GGFGLG 83

RESULT 3  
 ID 002402 PRELIMINARY; PRT; 738 AA.  
 AC 002402;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE INSOLUBLE PROTEIN.

OS Pinctada fucata.  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;  
 OC Pterioidea; Pterioidea; Pterioidea.  
 OX NCBI\_TaxID=50426;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97320490; PubMed=9177341;  
 RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,  
 RA Nakashima K., Takahashi T.;  
 RT "Structures of mollusc shell framework proteins.";  
 RL Nature 387:563-564(1997).  
 DR EMBL: D86074; BAA20466.1; -;  
 SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 11.9%; Score 8; DB 5; Length 738;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 493 GGFGLG 500

RESULT 4  
 ID 076271 PRELIMINARY; PRT; 904 AA.  
 AC 076271;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE NONGRADIENT BYSSAL PRECURSOR.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98393676; PubMed=9724735;  
 RA Qin X.X., Waite J.H.;  
 RT "A potential mediator of collagenous block copolymer gradients in  
 mussel byssal threads.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10517-10522(1998).  
 DR EMBL: AF043944; AAC33847.1; -;  
 DR InterPro: IPR000087; -;  
 DR Pfam: PF01391; Collagen: 7.  
 SQ SEQUENCE 904 AA; 77883 MW; 5529135651AD4C40 CRC64;

Query Match 11.9%; Score 8; DB 5; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 119 GGFGLG 126

RESULT 5  
 ID 044367 PRELIMINARY; PRT; 922 AA.  
 AC 044367;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PRECOLLAGEN D.  
 GN PRECOL-D.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-FOOT;  
 RX MEDLINE-98070444; PubMed-9405478;  
 RA Qin X.X., Coyne K.J., Waite J.H.;  
 RT "Tough tendons. Mussel byssus has collagen with silk-like domains.";  
 RL J. Biol. Chem. 273:32623-32627(1997).  
 DR EMBL: AF029249; AAB96638.1; -  
 DR InterPro: IPR000087; -  
 DR Pfam: PF01391; Collagen 7;  
 SQ SEQUENCE 922 AA; 80306 MW; 599D155E47A2C24A CRC64;

Query Match 11.9%; Score 8; DB 5; Length 922;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 132 GGFGLG 139

RESULT 6  
 ID 006810 PRELIMINARY; PRT; 1329 AA.  
 AC 006810;  
 DT 01-JUL-1997 (TRENBLREL. 04, Created)  
 DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)  
 DE HYPOTHELICAL 107.4 KDA PROTEIN.  
 GN GPRS-FAMILY OR RV1450C OR MTCY493.04.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires K., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z95844; CAB09271.1; -  
 DR Tuberculist: RV1450C; -  
 DR InterPro: IPR000084; -  
 DR STRAIN-CV. N-O-9;  
 DR InterPro: IPR002173; -  
 DR InterPro: IPR003536; -  
 DR Pfam: PF00934; PE; 1.  
 DR PRINTS: PRO1370; TRNSINTIMINR.  
 DR ProDom: PD001223; -; 1.  
 DR PROSITE: PS00583; PFKR\_KINASES\_1; UNKNOWN\_2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1329 AA; 107410 MW; 777A125F6DBAB234 CRC64;

Query Match 11.9%; Score 8; DB 2; Length 1329;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 Db 1318 GGFGLG 1325

RESULT 7  
 ID 09SD66 PRELIMINARY; PRT; 100 AA.

AC Q9SD66;  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)  
 DE HYPOTHELICAL 10.5 KDA PROTEIN.  
 GN F3112.120.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choisme N., Robert C., Brotier P., Wincker P., Catolico L.,  
 RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Queller F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL133292; CAB61953.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 100 AA; 10530 MW; 9D98F9F0B5EDBAD CRC64;

Query Match 10.4%; Score 7; DB 10; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 7  
 Db 89 GGFGLG 95

RESULT 8  
 ID Q9LEP7 PRELIMINARY; PRT; 107 AA.  
 AC Q9LEP7;  
 DT 01-OCT-2000 (TRENBLREL. 15, Created)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)  
 DE PUTATIVE GLYCINE-RICH PROTEIN.  
 OS Brassica napus (Rapeseed).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. N-O-9;  
 RA Bowers N.L., Trick M.;  
 RT "Microsynteny at the FCA region between Arabidopsis thaliana and  
 RT Brassica napus.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ293726; CAC01931.1; -  
 DR InterPro: IPR001525; -  
 DR PROSITE: PS00095; C5\_MTASE\_2; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 107 AA; 9714 MW; 0A7BE4A01B09B1B3 CRC64;

Query Match 10.4%; Score 7; DB 10; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGFGLG 8  
 Db 67 GGFGLG 73

RESULT 9  
 ID Q21073 PRELIMINARY; PRT; 284 AA.

AC Q21073;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
 DE K01A6.4 PROTEIN.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Coltage A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner M., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lighting J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RT Nature 368:32-38(1994).  
 RL EMBL: Z68750; CA92963.1; -;  
 DR SEQUENCE 284 AA; 27053 MW; C5A748DBE72DE82D CRC64;  
 SO

Query Match 10.4%; Score 7; DB 5; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GFGGGLG 7  
 |||||  
 Db 259 GFGGGLG 265

RESULT 10  
 Q9NL74 PRELIMINARY; PRT; 313 AA.  
 ID Q9NL74;  
 AC Q9NL74;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE DIHYDROROTATE DEHYDROGENASE.  
 GN PYR4.  
 OS Leishmania mexicana amazonensis.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=31284;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99096912; PubMed-9878395;  
 RX Gao G., Nara T., N-Shimada J., Aoki T.;  
 RA "Novel organization and sequences of five genes encoding all six  
 RT enzymes for de novo pyrimidine biosynthesis in Trypanosoma cruzi";  
 RL J. Mol. Biol. 285:149-161(1999).  
 DR EMBL: AB029444; BAA94299.1; -;  
 DR InterPro: IPR000408; -;  
 DR InterPro: IPR001295; -;  
 DR InterPro: IPR003009; -;  
 DR Pfam: PF01180; DHODHase; 1.  
 DR PROSITE: PS00626; RCCL\_2; UNKNOWN.1.  
 SO SEQUENCE 313 AA; 33697 MW; EB7558522D1E8E0A CRC64;  
 Query Match 10.4%; Score 7; DB 5; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 GFGGGLG 8  
 |||||  
 Db 217 GFGGGLG 223

RESULT 11  
 ID 076140 PRELIMINARY; PRT; 314 AA.  
 AC 076140;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE DIHYDROROTATE OXIDASE.  
 GN DHO OXIDASE OR PYR4.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 NCBI\_TaxID=5693;  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-TULAHUEN;  
 RA Gao G., Nara T., Nakajima-Shimada J., Aoki T.;  
 RT "Pyrimidine-genes in Trypanosoma";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-TULAHUEN;  
 RX MEDLINE-99096912; PubMed-9878395;  
 RA Gao G., Nara T., Nakajima-Shimada J., Aoki T.;  
 RT "Novel organization and sequences of five genes encoding all six  
 RT enzymes for de novo pyrimidine biosynthesis in Trypanosoma cruzi";  
 RL J. Mol. Biol. 285:149-161(1999).  
 DR EMBL: AB010286; BAA31360.1; -;  
 DR EMBL: AB017765; BAA74526.1; -;  
 DR HSSP: P54321; 2DOR.  
 DR InterPro: IPR001295; -;  
 DR InterPro: IPR003009; -;  
 DR Pfam: PF01180; DHODHase; 1.  
 SO SEQUENCE 314 AA; 34052 MW; 65163858A022BD80 CRC64;

Query Match 10.4%; Score 7; DB 5; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGGLG 8  
 |||||  
 Db 218 GFGGGLG 224

RESULT 12  
 ID 09RV66 PRELIMINARY; PRT; 438 AA.  
 AC 09RV66;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE GTP-BINDING PROTEIN ORG.  
 GN DR0084.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 NCBI\_TaxID=1299;  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RI;  
 RX MEDLINE-20036896; PubMed-10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.;



RA Fraser C.M.;  
 RT Genome sequence of the radioresistant bacterium Deinococcus  
 RL radiodurans R1.";  
 DR Science 286:1571-1577(1999).  
 DR EMBL: AE001871; AAF09676.1; -  
 DR TIGR: DR0084; -  
 DR InterPro: IPR000765; -  
 DR InterPro: IPR001684; -  
 DR Pfam: PF01018; GTP1\_OBG; 1.  
 DR PRINTS: PR00326; GTP1\_OBG.  
 DR PRODOM: PD003114; -; 1.  
 DR PROSITE: PS00905; GTP1\_OBG; 1.  
 SO SEQUENCE 438 AA; 47836 MW; D0DC5612543F2989 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGGRG 10  
 |||||  
 DB 120 GGLGGRG 126

RESULT 13

083238 PRELIMINARY; PRT; 450 AA.  
 AC 083238;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PREPROMEIN TRANSLOCASE SUBUNIT (SECY).  
 GN TP0208.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 DR EMBL: AE001203; AAC65198.1; -  
 DR TIGR: TP0208; -  
 DR InterPro: IPR001220; -  
 DR InterPro: IPR001525; -  
 DR InterPro: IPR002208; -  
 DR Pfam: PF00344; secy; 1.  
 DR PRINTS: PR00303; SECYTRNLASE.  
 DR PROSITE: PS00095; C5\_MWASE.2; UNKNOWN\_1.  
 DR PROSITE: PS00307; LECTIN\_LEGOME\_BETA; UNKNOWN\_1.  
 DR PROSITE: PS00756; SECY\_2; UNKNOWN\_1.  
 SO SEQUENCE 450 AA; 50265 MW; 892BF3217F3E6E7E CRC64;

Query Match 10.4%; Score 7; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GGLGGRG 11  
 |||||  
 DB 427 GGLGGRG 433

RESULT 14

010787 PRELIMINARY; PRT; 463 AA.  
 ID 010787;  
 AC 010787;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HYPOTHETICAL 49.1 KDA PROTEIN CY48.28C.  
 GN RV1537 OR MTCY48.28C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO UMCC, SAMB, AND MUCB PROTEINS.  
 DR EMBL: Z74020; CAA98325.1; -  
 DR Tuberculist; RV1537; -  
 DR InterPro: IPR001126; -  
 DR Pfam: PF00817; IMS; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 463 AA; 49107 MW; 43085A44CE8CC637 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGGRG 10  
 |||||  
 DB 29 GGLGGRG 35

RESULT 15  
 050943 PRELIMINARY; PRT; 465 AA.  
 ID 050943;  
 AC 050943;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 52.8 KDA PROTEIN.  
 GN BBA50.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OG Plasmid lp54.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Utterback T., Wathey L., McDonald L., Artach P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
 RT burgdorferi.";  
 RL Nature 390:580-586(1997).  
 DR EMBL: AE000790; AAC66282.1; -  
 DR TIGR: BBA50; -  
 KW Hypothetical protein; Plasmid.  
 SO SEQUENCE 465 AA; 52773 MW; 8223FBE38B7E5B7 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGAG 8  
 |||||  
 DB 44 GFGGAG 50

## RESULT 16

ID 09L8K4 PRELIMINARY; PRT; 556 AA.  
 AC 09L8K4;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PE-PGRS HOMOLOG MAG24-2.  
 GN MAG24-2.  
 OS Mycobacterium marinum.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M:  
 RA Ramakrishnan L., Federspiel N.A., Falkow S.;  
 RT "A Mycobacterial Protein of the Repetitive Glycine-rich PE-PGRS Family  
 Is Required for Macrophage Replication and Contributes to Virulence."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF201682; AAF65169.1;  
 DR InterPro: IPR000084;  
 DR InterPro: IPR001412;  
 DR Pfam: PF00934; PE; 1.  
 DR ProDom: PD001223;  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
 SO SEQUENCE 556 AA; 54576 MW; 8C8129B5D368A634 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 556;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GFGGAG 10  
 |||||  
 DB 181 GFGGAG 187

## RESULT 17

ID 09FU14 PRELIMINARY; PRT; 620 AA.  
 AC 09FU14;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PUTATIVE RECEPTOR KINASE.  
 GN P06986G3.13.  
 OS Oryza sativa (Rice)  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
 OC Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: p06986G3."  
 DE Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP002747; BAB17332.1;  
 KW Receptor; Kinase.  
 SO SEQUENCE 620 AA; 69544 MW; C95560BF0BC1ABC6 CRC64;

Query Match 10.4%; Score 7; DB 10; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 GGRKPS 14  
 |||||  
 DB 30 GGRKPS 36

## RESULT 18

ID 068071 PRELIMINARY; PRT; 652 AA.  
 AC 068071;  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE DNA GYRASE SUBUNIT B (Ec 5.99.1.3).  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
 OC Rhodospirillum.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SB1003;  
 RX MEDLINE-97404404; PubMed-9256491;  
 RA Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Fomstein M.;  
 RT "Sequence of a 189-kb segment of the chromosome of Rhodospirillum  
 rubrum capsulatus SB1003."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
 OF DOUBLE-STRANDED DNA.  
 CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.  
 CC EMBL: AF010496; AAC16157.1;  
 DR HSP: P06982; 1A76.  
 DR InterPro: IPR000410;  
 DR InterPro: IPR001241;  
 DR InterPro: IPR002288;  
 DR InterPro: IPR002936;  
 DR Pfam: PF00204; DNA\_topoisomII; 1.  
 DR Pfam: PF00986; DNA\_gyraseB\_C; 1.  
 DR Pfam: PF01751; Toprim; 1.  
 DR PRINTS: PR00418; TP12FAMILY.  
 DR ProDom: PD000616;  
 DR ProDom: PD149633;  
 DR PROSITE: PS00177; TOPOISOMERASE\_II; 1.  
 DR SMART: SM00433; TOP2c; 1.  
 KW ATP-binding; Isomerase; Topoisomerase.  
 SO SEQUENCE 652 AA; 71524 MW; 4CE2687B47F4289D CRC64;

Query Match 10.4%; Score 7; DB 2; Length 652;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GIGGRK 11  
 |||||  
 DB 573 GIGGRK 579

## RESULT 19

ID 022537 PRELIMINARY; PRT; 682 AA.  
 AC 022537;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE SIMILARITY TO COLLAGEN.  
 GN T17H7.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,  
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RT Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Favello A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U42841; AAC48170.1; -;  
 SO SEQUENCE 682 AA; 72150 MW; 73933507D23E98A2 CRC64;

Query Match 10.4%; Score 7; DB 5; Length 682;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGLGGRG 10  
 |||||  
 DB 587 GGLGGRG 593

RESULT 20  
 Q9H6R3 PRELIMINARY; PRT; 686 AA.  
 AC Q9H6R3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CDNA: FLJ21963 FIS, CLONE HEP05563.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Okeyashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK025616; BAB15190.1; -;  
 SO SEQUENCE 686 AA; 74778 MW; 3386338C6FFFD7E8 CRC64;

Query Match 10.4%; Score 7; DB 4; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGLGGRG 10  
 |||||  
 DB 44 GGLGGRG 50

RESULT 21  
 087472 PRELIMINARY; PRT; 753 AA.  
 AC 087472;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE TPRJ.  
 OS Treponema pallidum (subsp. pertenue) (Yaws treponeme).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI\_TaxID=168;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stamm L.V., Greene S.R., Bergen H.L., Hardham J.M., Barnes N.Y.;  
 RT "Identification and Sequence Analysis of Treponema pallidum tprJ, a  
 RT FEMS Microbiol. Lett. 0:0-0(1998).  
 RL EMBL: AF073527; AAC83339.1; -;  
 SO SEQUENCE 753 AA; 80692 MW; BB6C99841BDF6A8 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 753;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGGLGG 8  
 |||||  
 DB 166 GFGGLGG 172

RESULT 22  
 083337 PRELIMINARY; PRT; 756 AA.  
 AC 083337;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DE TPR PROTEIN G (TPRG).  
 GN TP0317.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=9833270; PubMed=9655876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.R., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,  
 RA McDonald L., Atlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."  
 RL Science 281:375-388(1998).  
 DR EMBL: AE001211; AAC65301.1; -;  
 TIGR: TP0317; -;  
 SO SEQUENCE 756 AA; 81324 MW; 4D1F6B8CE51738A CRC64;

Query Match 10.4%; Score 7; DB 2; Length 756;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGGLGG 8  
 |||||  
 DB 169 GFGGLGG 175

RESULT 23  
 007894 PRELIMINARY; PRT; 758 AA.  
 AC 007894;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE MAJOR OUTER SHEATH PROTEIN HOMOLOG MSP.  
 GN MSP OR TP0621.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RA Hardham J.M.;  
 RL Thesis (1995), Unknown Institution.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RA Stamm L.V., Barnes N.Y., Frye J.F.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RA MEDLINE=9833770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Ariach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sanusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 DR EMBL; U88957; AAC83171.1; -  
 DR EMBL; AE001237; AAC65595.1; -  
 DR TIGR; TP0621; -  
 SQ SEQUENCE 758 AA; 81369 MW; FFF93A8951E261B8 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLG 8  
 |||||  
 DB 169 GFGGLG 175

RESULT 24  
 O9KHF2 PRELIMINARY; PRT; 758 AA.  
 AC O9KHF2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE TPRJ.  
 OS *Treponema pallidum*.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; *Treponema*.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SS14;  
 RA Stamm L.V., Bergen H.L.;  
 RT "Intrastain Heterogeneity of the TprK Protein of the *Treponema*  
 RT *pallidum* Nichols Strain and Street Strain 14.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF261075; AAF76884.1; -  
 SQ SEQUENCE 758 AA; 81334 MW; DF91F1ED0F6CB3D7 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLG 8  
 |||||  
 DB 169 GFGGLG 175

RESULT 25  
 O9KHF1 PRELIMINARY; PRT; 758 AA.  
 ID O9KHF1;  
 AC O9KHF1;

DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE TPRJ.  
 OS *Treponema pallidum*.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; *Treponema*.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SS14;  
 RA Stamm L.V., Bergen H.L.;  
 RT "Intrastain Heterogeneity of the TprK Protein of the *Treponema*  
 RT *pallidum* Nichols Strain and Street Strain 14.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF261076; AAF76885.1; -  
 SQ SEQUENCE 758 AA; 81285 MW; E4B882007BBF321B CRC64;

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

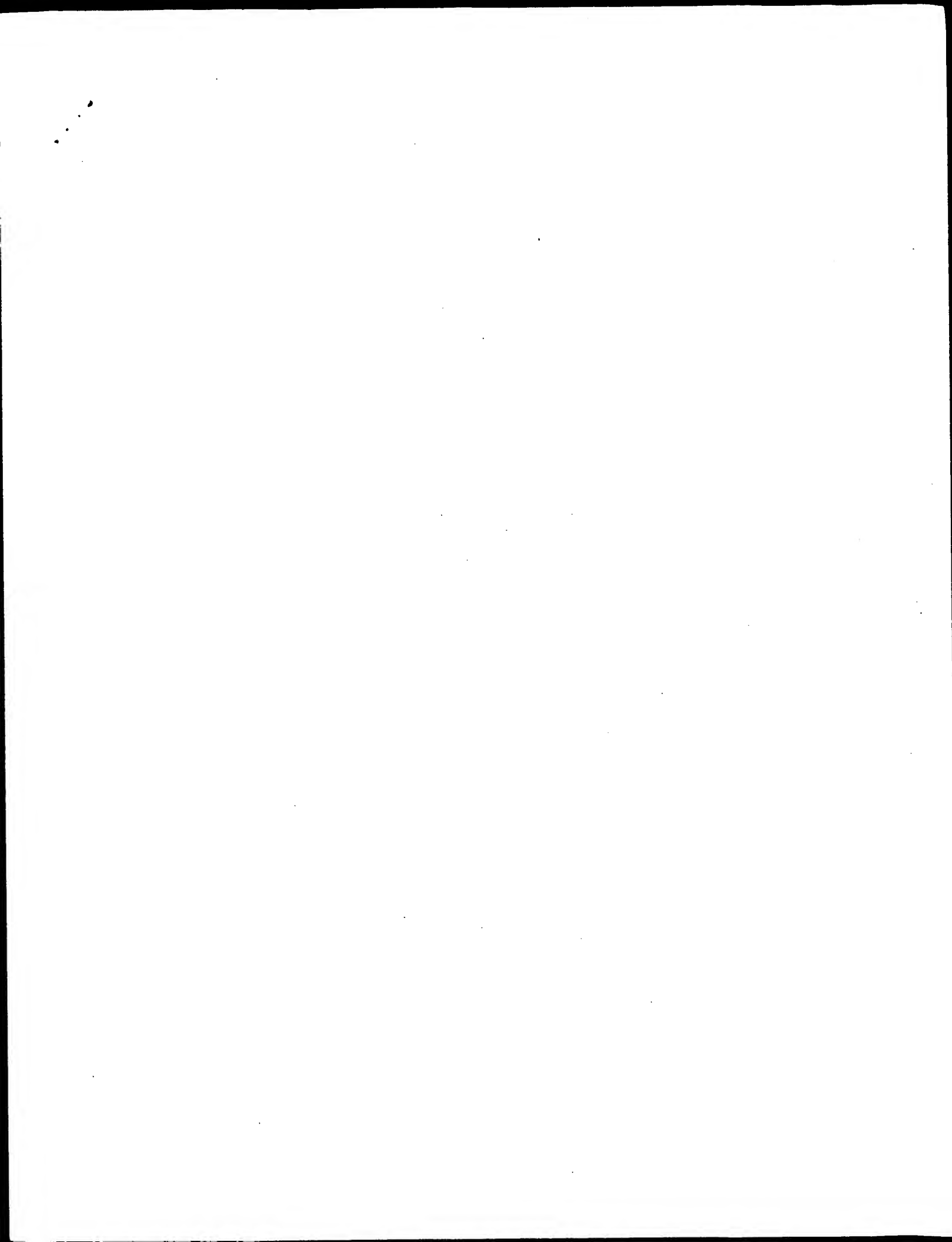
OY 2 GFGGLG 8  
 |||||  
 DB 169 GFGGLG 175

Search completed: July 12, 2001, 06:22:54  
 Job time: 345 sec

Thu Jul 12 06:30:38 2001

us-09-506-978-1.oligo.rspt

---





DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2591.  
 GN RV2591 OR MTCY227.10C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoon K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Horsby T., Jags K., Krogh A., Mclean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z77724; CAB01283.1; -;  
 DR Tuberculist: RV2591; -;  
 DR InterPro: IPR000084; -;  
 DR Pfam: PF00934; Pf.1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 543 AA; 46287 MW; 59730339E5D2DF59 CRC64;

Query Match 11.9%; Score 8; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGGGLGG 8  
 DB 380 GGGGGLGG 387

RESULT 3  
 ID RUI7\_YEAST STANDARD; PRT; 300 AA.  
 AC 000916;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA HOMOLOG.  
 GN SNPI OR YI061C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=91330888; PubMed=1714384;  
 RA Smith V., Barrell B.G.;  
 RT "Cloning of a yeast U1 snRNP 70K protein homologue: functional  
 RT conservation of an RNA-binding domain between humans and yeast.";  
 RL EMBL J. 10:2627-2634(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;

RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jags K., Jones M.,  
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1. IT BINDS  
 CC STEW LOOP I OF U1 snRNA.  
 CC -----  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z38060; CAA86162.1; -;  
 DR EMBL: X59986; CAA42602.1; -;  
 DR PIR: S16815; S16815.  
 DR PIR: S48418; S48418.  
 DR HSP: P19339; 2SXL.  
 DR TRANSFAC: T01242; -;  
 DR SGD: S0001323; SNPI.  
 DR InterPro: IPR000504; -;  
 DR Pfam: PF00076; Rrm.1.  
 DR PROSITE: PS50102; RRM.1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; 1.  
 DR Nucleic acid binding protein; RNA-binding.  
 FT Nucleic acid binding protein; RNA-binding (RRM).  
 SQ SEQUENCE 300 AA; 34447 MW; 7D9E47BE6FE1EB8 CRC64;

Query Match 10.4%; Score 7; DB 1; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGGRG 10  
 DB 201 GGLGGRG 207

RESULT 4  
 ID DHSO\_SCHPO STANDARD; PRT; 360 AA.  
 AC P36624;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PUTATIVE SORBITOL DEHYDROGENASE (EC 1.1.1.14) (L-IDITOL  
 DE 2-DEHYDROGENASE) (TMS1 PROTEIN).  
 GN TMS1 OR SPC1773.05C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Lyne M., Rajandream M.A., Barrell B.G., Beck A., Borzym K., Klages S.,  
 RA Langer I., Reinhardt R.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 5-360 FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=94039112; PubMed=8223615;  
 RA Wagner P., Grimaldi M., Jenkins J.R.;  
 RT "Putative dehydrogenase tms1 suppresses growth arrest induced by a  
 RT p53 tumour mutant in fission yeast.";  
 RL Eur. J. Biochem. 217:731-736(1993).



```

CC -1- FUNCTION: SUPPRESSES GROWTH ARREST INDUCED BY A P53 TUMOR MUTANT
CC IN FISSION YEAST.
CC -1- CATALYTIC ACTIVITY: L-SORBITOL + NAD(+) = L-SORBOSE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL033389; CA21910.1; -
DR EMBL: X74422; CA852443.1; ALT_INIT.
DR PIR: S35981; S35981.
DR PIR: S38345; S38345.
DR HSSP: P07846; 1SDG.
DR InterPro: IPR002085; -
DR InterPro: IPR002328; -
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC.1.
DR KW Oxidoreductase; Zinc; NAD.
FT METAL 42 42 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 67 67 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 153 153 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 360 AA; 38851 MW; ADEB3DDBA165224C CRC64;

Query Match 10.4%; Score 7; DB 1; Length 360;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 PGCVRL 50
DB 93 PGCVRL 99

RESULT 5
DNJ2_MYCLE STANDARD; PRT; 378 AA.
AC 049762;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ2.
OS DNAJ2 OR B1937_F2-56.
SN Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodison K., Smith D.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -1- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -1- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: U00016; AAA17167.1; -
DR HSSP: P25685; 1HDJ.
DR InterPro: IPR001305; -
DR InterPro: IPR001623; -
DR InterPro: IPR002939; -
DR InterPro: IPR003095; -
DR Pfam: PF00226; DnaJ_1.
DR Pfam: PF01556; DnaJ_C_1.
DR Pfam: PF00684; DnaJ_CXXCXXG.1.
DR PRINTS: PR00625; DNAJPROTEIN.
DR PROSITE: PS00636; DNAJ_1; FALSE_NEG.
DR PROSITE: PS50076; DNAJ_2; 1.
DR PROSITE: PS00637; DNAJ_CXXCXXG; FALSE_NEG.
KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.
FT DOMAIN 4 68 J-DOMAIN.
FT REPEAT 80 111 GLY-RICH.
FT REPEAT 141 148 CXXCXXG MOTIF.
FT REPEAT 158 165 CXXCXXG MOTIF.
FT REPEAT 184 191 CXXCXXG MOTIF.
FT REPEAT 198 205 CXXCXXG MOTIF.
FT METAL 141 141 ZINC 1 (BY SIMILARITY).
FT METAL 144 144 ZINC 1 (BY SIMILARITY).
FT METAL 158 158 ZINC 2 (BY SIMILARITY).
FT METAL 161 161 ZINC 2 (BY SIMILARITY).
FT METAL 184 184 ZINC 2 (BY SIMILARITY).
FT METAL 187 187 ZINC 2 (BY SIMILARITY).
FT METAL 198 198 ZINC 1 (BY SIMILARITY).
FT METAL 201 201 ZINC 1 (BY SIMILARITY).
SQ SEQUENCE 378 AA; 40399 MW; 73DD8BFFCF62153 CRC64;

```

```

Query Match 10.4%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GFGGGLG 7
DB 82 GFGGGLG 88

RESULT 6
DNJ2_MYCTU STANDARD; PRT; 382 AA.
AC 005825;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ2.
OS DNAJ2 OR RV2373C OR MTCY27.07.
SN Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RL Nature 393:537-544 (1998).
CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -1- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



RX	MEDLINE=94214498; PubMed=7512862.
RA	Reis A, Hennies H.-C., Langbein L., Digweed M., Mischke D.,
RA	Dreschler M., Schroek E., Royer-Pokora B., Franke W.W., Spertling K,
RT	Kuester W.;
RT	"Keratins 9 gene mutations in epidermolytic palmoplantar keratoderma
RT	(EPPK).";
RL	Nat. Genet. 6:174-179(1994).
RM	[4]
RP	VARIANTS: EPPK TRP-162 AND SER-167.
RX	MEDLINE=95164983; PubMed=7532199;
RA	Rothnagel J.A., Wojcik S., Liefer K.M., Dominey A.M., Huber M.,
RA	Hohl D., Roop D.R.;
RT	"Mutations in the 1A domain of Keratin 9 in patients with
RT	epidermolytic palmoplantar keratoderma.";
RL	J. Invest. Dermatol. 104:430-433(1995).
CC	-1- FUNCTION: MAY SERVE AN IMPORTANT SPECIAL FUNCTION EITHER IN THE
CC	MATURE PALMAR AND PLANTAR SKIN TISSUE OR IN THE MORPHOGENIC
CC	PROGRAM OF THE FORMATION OF THESE TISSUE.
CC	-1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN THE TERMINALLY DIFFERENTIATED
CC	EPIDERMIS OF PALMS AND SOLES.
CC	-1- DISEASE: DEFECTS IN KR9 ARE A CAUSE OF EPIDERMOLYTIC PALMOPPLANTAR
CC	KERATODEMA (EPPK), AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY
CC	DIFFUSE THICKENING OF THE EPIDERMIS ON THE ENTIRE SURFACE OF
CC	PALMS AND SOLES SHARPLY BORDERED WITH ERYTHEMATOUS MARGINS.
CC	-1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC	MICROFILAMENTULAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC	(NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC	-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC	-1- CAUTION: WAS ORIGINALLY (REF. 2) THOUGHT TO BE A 60 KDA CHAIN OF
CC	PLACENTAL SCATTER PROTEIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way used
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X75015; CAA52924.1; -;
DR	EMBL; Z29074; CAA82315.1; -;
DR	EMBL; S69510; AAC60619.1; -;
DR	PIR; B35494; B35494.
DR	HSP; P02876; ZWGC.
DR	MIU; 144200; -;
DR	InterPro; IPR001664; -;
DR	InterPro; IPR002957; -;
DR	Pfam; PF00038; filament; 1.
DR	PRINTS; PR01248; TYPE1KERATIN.
DR	PROSITE; PS00226; IF; 1.
KW	Intermediate filament; Coiled coll; Heptad repeat pattern; Keratin;
KW	Disease mutation.
FT	DOMAIN
FT	1..151 HEAD.
FT	152..460 ROD.
FT	461..622 TAIL.
FT	DOMAIN
FT	152..187 COIL 1A.
FT	DOMAIN
FT	188..206 LINKER 1.
FT	DOMAIN
FT	207..298 COIL 1B.
FT	DOMAIN
FT	299..321 LINKER 12.
FT	DOMAIN
FT	322..460 COLL 2.
FT	DOMAIN
FT	14..20 POLY-GLY.
FT	VARIANT
FT	160..160 N -> R (IN EPPK).
FT	/FTID-VAR_003822.
FT	VARIANT
FT	162..162 R -> Q (IN EPPK).
FT	/FTID-VAR_003823.
FT	VARIANT
FT	162..162 R -> W (IN EPPK).
FT	/FTID-VAR_003824.
FT	VARIANT
FT	167..167 L -> S (IN EPPK).
FT	/FTID-VAR_003825.
FT	CONFLICT
FT	12..12 T -> SR (IN CAA82315).
QO	SEQUENCE 622 AA; 61987 MW; 89BC3825D4B5ED94 CRC64;

Query Match	10.4%;	Score 7;	DB 1;	Length 622;
Best Local Similarity	100.0%;	Pred. No. 5.4;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
				0
Qy	2	GGGGLG 8		
Db	129	GGGGLG 135		
RESULT	9			
ID	SCXP_ANDMA	STANDARD;	PRT;	35 AA.
AC	P01498;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	NEUROTOXIN P2.			
OS	Androctonus mauretanicus mauretanicus (Scorpion).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;			
OX	NCBI_TaxID=6860;			
RN	[1]			
RP	SEQUENCE.			
RX	MEDLINE=85193276; PubMed=3992595;			
RA	Rosso J.P., Rochat H.;			
RT	"Characterization of ten proteins from the venom of the Moroccan			
RT	scorpion Androctonus mauretanicus mauretanicus, six of which are			
RT	toxic to the mouse.";			
RL	Toxicon 23:113-125(1985).			
CC	-I- SIMILARITY: BELONGS TO THE SHORT SCORPION TOXIN FAMILY.			
DR	PIR, A01758; NTSRPM.			
DR	HSSP; P15222; 1S1S.			
KM	Neurotoxin.			
FT	DISULFID 1 18 BY SIMILARITY.			
FT	DISULFID 4 25 BY SIMILARITY.			
FT	DISULFID 15 30 BY SIMILARITY.			
FT	DISULFID 19 32 BY SIMILARITY.			
SO	SEQUENCE 35 AA; 3673 MW; 213EG926289EB5A CRC64;			
Query Match	9.0%;	Score 6;	DB 1;	Length 35;
Best Local Similarity	100.0%;	Pred. No. 6.8;		
Matches	6;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
				0.
Qy	7	GGRGKC 12		
Db	20	GGRGKC 25		
RESULT	10			
ID	YMC3_OENBE	STANDARD;	PRT;	79 AA.
AC	P08747;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	HYPOTHETICAL 9.2 KDA PROTEIN IN COXIII REGION.			
OS	Oenothera lamarckiana (Bertero's evening primrose).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Myrtales; Onagraceae; Oenothera.			
OX	NCBI_TaxID=3950;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV. MONZIA;			
RA	Hiesel R., Schobel W., Schuster W., Brennicke A.;			
RT	"The cytochrome oxidase subunit I and subunit III genes in Oenothera			
RT	mitochondria are transcribed from identical promoter sequences.";			
RL	EMBO J. 6:29-34(1987).			
CC	-I- MISCELLANEOUS: POSITION 54 IS MODIFIED BY RNA EDITING.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04764; CAA28457.1; ALT\_SEQ.  
 KW Microchondion: RNA editing: Hypothetical protein.  
 SQ SEQUENCE 79 AA: 9196 MW; DDE115BC7E3718CC CRC64;

Query Match 9.0%; Score 6; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 KVCVPR 63  
 |||||  
 DB 35 KVCVPR 40

RESULT 11  
 MIFH\_WUCBA STANDARD; PRT; 114 AA.  
 ID MIFH\_WUCBA 044786;

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE MACROPHAGE MIGRATION INHIBITORY FACTOR HOMOLOG.  
 GN MIF.

OS Wuchereria bancrofti.  
 CC Eukaryota; Metazoa; Chromadorea; Splirurida; Filarioidae;  
 CC Onchocercidae; Wuchereria.  
 OX NCBI\_TaxID=62933;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-BRAZIL;  
 RA Scott A.L., Josh S., Pastrana D., Eisinger S.W., Marson A.,  
 RA Ragshavan N.,  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MIF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AF040629; AAC82615.1; -  
 DR HSSP: P14174; MIF.  
 DR InterPro: IPR001398; -  
 DR Pfam: PF01187; MIF; 1.  
 DR PROSITE: PS01158; MIF; 1.  
 KM Cytokine.  
 FT INT\_MEF.  
 SQ SEQUENCE 114 AA: 12233 MW; E3F2A89F17BF1E3D CRC64;

Query Match 9.0%; Score 6; DB 1; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 PNVVPR 35  
 |||||  
 DB 23 PNVVPR 28

RESULT 12  
 RL3\_PLARQ STANDARD; PRT; 140 AA.  
 ID RL3\_PLARQ

AC P72233;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 50S RIBOSOMAL PROTEIN L3 (FRAGMENT).  
 GN RPLC.

OS Planobispora rosea.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Streptosporangineae; Streptosporangaceae;  
 CC Planobispora.  
 OX NCBI\_TaxID=35762;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 53733;  
 RX MEDLINE=97055420; PubMed=8899707;  
 RA Sostio M., Amat G., Cappellano C., Sarubbi E., Monti F.,  
 RA Donadio S.;  
 RT "An elongation factor Tu (EF-Tu) resistant to the EF-Tu inhibitor  
 RT GE2270 in the producing organism Planobispora rosea.";  
 RL Mol. Microbiol. 22:43-51(1996).

CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY  
 CC PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF  
 CC THE RIBOSOME (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X98830; CAA67347.1; -  
 DR InterPro: IPR000597; -  
 DR Pfam: PF00297; Ribosomal\_L3; 1.  
 DR PROSITE: PS00474; RIBOSOMAL\_L3; 1.  
 KM Ribosomal protein; rRNA-binding.  
 FT NON TER 140  
 SQ SEQUENCE 140 AA: 15232 MW; B50630F7A0A4849D CRC64;

Query Match 9.0%; Score 6; DB 1; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGLG 7  
 |||||  
 DB 130 GFGLG 135

RESULT 13  
 PA24\_RAT STANDARD; PRT; 150 AA.  
 ID PA24\_RAT P39878;

DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE 14 KDa PHOSPHOLIPASE A2 PRECURSOR (EC 3.1.1.4) (PHOSPHATIDYLCHOLINE  
 DE 2-ACYLHYDROLASE) (PLA2-8) (GROUP IIC PHOSPHOLIPASE A2).  
 GN PLA2G2C.

OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-Brain;  
 RX MEDLINE=94364995; PubMed=8083202;  
 RA Chen J., Engle S.J., Seilhamer J.J., Tischfield J.A.;  
 RT "Cloning and characterization of novel rat and mouse low molecular  
 RT weight Ca(2+)-dependent phospholipase A2s containing 16 cysteines.";  
 RL J. Biol. Chem. 269:23018-23024(1994).

CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-  
CC ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.  
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERYL-  
CC PHOSPHOCHOLINE + A FATTY ACID ANION.  
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. MAY REPRESENT  
CC A NEW SUBTYPE OF GROUP II PLA2.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: U07798; AAA57473.1; ALT\_INIT.  
CC DR HSSP: P14418; 1PSJ.  
CC DR InterPro: IPR001211; -.  
CC DR Pfam: PF00068; Phoslip; 1.  
CC DR PROSITE: PS00118; PA2\_HIS; 1.  
CC DR PROSITE: PS00119; PA2\_ASP; 1.  
CC KM Hydrolyase; Lipid degradation; Signal; Calcium.  
CC FT SIGNAL 1 20  
CC FT CHAIN 21 150  
CC FT ACT\_SITE 67 67 BY SIMILARITY.  
CC FT CA\_BIND 68 68 POTENTIAL.  
CC FT ACT\_SITE 115 115 BY SIMILARITY.  
CC FT DISULFID 46 143 BY SIMILARITY.  
CC FT DISULFID 48 64 BY SIMILARITY.  
CC FT DISULFID 63 121 BY SIMILARITY.  
CC FT DISULFID 69 150 BY SIMILARITY.  
CC FT DISULFID 70 114 BY SIMILARITY.  
CC FT DISULFID 79 107 BY SIMILARITY.  
CC FT DISULFID 97 112 BY SIMILARITY.  
CC FT DISULFID 99 105 POTENTIAL.  
CC FT CARBOHYD 92 92 N-LINKED (GLCNAC...) (POTENTIAL).  
CC SQ SEQUENCE 150 AA: 16919 MW: F80575BE19A3BF64;  
  
Query Match 9.0%; Score 6; DB 1; Length 150;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 GLGGRG 10  
Db 49 GLGGRG 54  
  
RESULT 14  
MAUM\_METEX STANDARD; PRT; 220 AA.  
ID MAUM\_METEX  
AC Q49130;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE METHYLAMINE UTILIZATION FERREDOXIN-TYPE PROTEIN MAUM.  
GN MAUM.  
OS Methylobacterium extorquens.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Methylobacterium group; Methylobacterium.  
OX NCBI\_TaxID=408;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-AM1 / NCIB 9133;  
RX MEDLINE=84292425; PubMed=8021187;  
RA Chistoserdov A.Y., Chistoserdova L.V., McIntire W.S., Lidstrom M.E.;  
RT "Genetic organization of the mau gene cluster in Methylobacterium  
RT extorquens AM1: complete nucleotide sequence and generation and  
RT characteristics of mau mutants."  
RL J. Bacteriol. 176:4052-4065(1994).  
CC -1- FUNCTION: INVOLVED IN ELECTRON TRANSFER (POTENTIAL).  
CC -1- PATHWAY: METHYLAMINE UTILIZATION.  
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF

CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: L26406; AAB46941.1; -.  
CC DR HSSP: P46797; 1VTM.  
CC DR InterPro: IPR001450; -.  
CC DR Pfam: PF00037; Icd4; 2.  
CC DR PROSITE: PS00198; 4FE4S\_FERREDOXIN; 1.  
CC KM Electron transport; Iron-sulfur; 4Fe-4S.  
CC FT METAL 60 60 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 63 63 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 66 66 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 70 70 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 98 98 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 101 101 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 106 106 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 110 110 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 138 138 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 146 146 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 149 149 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 153 153 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 182 182 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 185 185 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 188 188 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
CC FT METAL 192 192 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
CC SQ SEQUENCE 220 AA: 23308 MW: EDB3CF81BE2947D9 CRC64;  
  
Query Match 9.0%; Score 6; DB 1; Length 220;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 GLGGRG 10  
Db 207 GLGGRG 212  
  
RESULT 15  
SURE\_HAEIN STANDARD; PRT; 249 AA.  
ID SURE\_HAEIN  
AC P45681;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE STATIONARY-PHASE SURVIVAL PROTEIN SURE HOMOLOG.  
GN SURE OR H10702.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-RD / RW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Kerslavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Kierlavage A.R., Bult C.J., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Georgiann N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).

```

CC -1- FUNCTION: NOT KNOWN. COULD BE A PHOSPHATASE.
CC -1- SIMILARITY: BELONGS TO THE SURE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32753; AAC2361.1; .
DR TIGR: H10702; .
DR InterPro: IPR002828; .
DR Pfam: PF01975; Sure; 1.
SQ SEQUENCE 249 AA; 27340 MW; D10F280C95266757 CRC64;

Query Match
Best Local Similarity 9.0%; Score 6; DB 1; Length 249;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VCRGCG 52
DB 172 VCRGCG 177

RESULT 16
RS2_MOUSE STANDARD; PRT; 293 AA.
AC P25444;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 40S RIBOSOMAL PROTEIN S2 (S4) (L1REP3 PROTEIN).
GN RPS2 OR RPS4 OR L1REP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88302198; PubMed=3405219;
RA Heller D.L., Gianola K.M., Leinward L.A.;
RT "A highly conserved mouse gene with a propensity to form pseudogenes
RT in mammals."
RL Mol. Cell. Biol. 8:2797-2803(1988).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Vauthey A.-L., Balroch A.;
RL Unpublished observations (JUN-1996).
CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 24 TO MODIFY THE N-TERMINUS SO AS TO
CC MAXIMIZE THE SIMILARITY WITH OTHER MEMBERS OF THIS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20632; AAA40074.1; ALT_FRAME.
DR PIR: A31139; A31139.
DR HSSP: P02357; 1PKP.
DR MGD: MGI:105110; L1rep3.
DR InterPro: IPR000851; .
DR Pfam: PF00333; Ribosomal_S5; 1.
DR PROSITE: PS00585; RIBOSOMAL_S5; 1.
DR Ribosomal protein, Repeat.
KW Ribosomal protein, Repeat.
SQ SEQUENCE 293 AA; 31217 MW; 90E5DB564AA624CC CRC64;

```

```

Query Match
Best Local Similarity 9.0%; Score 6; DB 1; Length 293;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10
DB 18 GLGGRG 23

RESULT 17
RS2_RAT STANDARD; PRT; 293 AA.
AC P27952;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 40S RIBOSOMAL PROTEIN S2.
GN RPS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 111-128.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=92041821; PubMed=1939063;
RA Suzuki K., Olivera J., Wool I.G.;
RT "Primary structure of rat ribosomal protein S2. A ribosomal protein
RT with arginine-glycine tandem repeats and RGGF motifs that are
RT associated with nucleolar localization and binding to ribonucleic
RT acids."
RL J. Biol. Chem. 266:20007-20010(1991).
CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57432; CAA40679.1; .
DR PIR: S18828; R3RT52.
DR HSSP: P02357; 1PKP.
DR InterPro: IPR000851; .
DR Pfam: PF00333; Ribosomal_S5; 1.
DR PROSITE: PS00585; RIBOSOMAL_S5; 1.
KW Ribosomal protein, Repeat.
FT DOMAIN 9 17 3 X 3 AA TANDEN REPEATS.
FT DOMAIN 22 29 2 X 4 AA TANDEN REPEATS.
FT DOMAIN 34 52 9 X 2 AA TANDEN REPEATS.
SQ SEQUENCE 293 AA; 31231 MW; 9092DB564AA624C9 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 293;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRG 10
DB 18 GLGGRG 23

RESULT 18
DNAL_DEIPR STANDARD; PRT; 307 AA.
ID DNAL_DEIPR
AC 034136;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)

```

DE CHAPERONE PROTEIN DNASJ (40 KDA HEAT SHOCK CHAPERONE PROTEIN) (HSP40).  
 GN DNASJ.  
 ON Deinococcus proteolyticus.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=55148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35074;  
 RX MEDLINE=97383250; PubMed=9236279;  
 RA Buxard K., Gupta R.S.;  
 RT "The sequences of heat shock protein 40 (DnaJ) homologs provide  
 RT evidence for a close evolutionary relationship between the  
 RT Deinococcus-thermus group and cyanobacteria.";  
 RT J. Mol. Evol. 45:193-205(1997).  
 RL -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNAB FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U93358; AAB96892.1; -  
 CC HSP: P08622; 1XBL.  
 DR InterPro: IPR001623; -  
 DR InterPro: IPR002939; -  
 DR Pfam: PF00226; DnaJ; 1.  
 DR Pfam: PF01556; DnaJ\_C; 1.  
 DR PROSITE: PS00636; DnaJ\_1; 1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 KW Chaperone; DNA replication; Heat shock.  
 FT DOMAIN 5  
 FT SEQUENCE 307 AA; 33244 MW; 9EE75066F2340F15 CRC64;

Query Match 9.0%; Score 6; DB 1; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FGGIGG 8  
 |||||

DB 132 FGGIGG 137

RESULT 19  
 ID KDCK\_HAEIN STANDARD; PRT; 314 AA.  
 AC P44452;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE 2-DEHYDRO-3-DEOXYGLUCONKINASE (EC 2.7.1.45) (2-KETO-3-  
 DE DEOXYGLUCONKINASE) (3-DEOXY-2-OXO-D-GLUCONATE KINASE) (KOG KINASE).  
 GN KOG OR H10049.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Gleyzer J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,

RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.U., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- CATALYTIC ACTIVITY: ATP + 2-DEHYDRO-3-DEOXY-D-GLUCONATE - ADP +  
 CC BETA-L-ARABINOSIDE 1-PHOSPHATE.  
 CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U32690; AAC21727.1; -  
 DR TIGR: H10049; -  
 DR InterPro: IPR002173; -  
 DR Pfam: PF00294; PKB; 1.  
 DR PROSITE: PS00583; PKB\_KINASES\_1; FALSE\_NEG.  
 DR PROSITE: PS00584; PKB\_KINASES\_2; 1.  
 KW Transferase; Kinase.  
 FT SEQUENCE 314 AA; 35462 MW; 24CA631C8B0485F9 CRC64;

Query Match 9.0%; Score 6; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 GYLNRK 56  
 |||||

DB 272 GYLNRK 277

RESULT 20  
 ID UL16\_HSV6U STANDARD; PRT; 335 AA.  
 AC P24442;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROTEIN U65.  
 DE U65 OR 11R.  
 GN Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).  
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseolovirus.  
 OX NCBI\_TaxID=10370;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90080132; PubMed=2152817;  
 RA Lawrence G.L., Chee M., Craxton M.A., Gompels U.A., Honess R.W.,  
 RA Barrell B.G.;  
 RT "Human herpesvirus 6 is closely related to human cytomegalovirus.";  
 RL J. Virol. 64:287-299(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95266321; PubMed=7747482;  
 RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
 RA Martin M.E., Efsthaliou S., Craxton M., Macaulay H.A.;  
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,  
 RT and genome evolution.";  
 RL Virology 209:29-51(1995).  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,  
 CC HSV-6 ORF11R, EBV-1 46, HCMV UL94, EBV BGFLF2, HSV 33, AND VZV 44.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



RN (1)  
RP SEQUENCE FROM



```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb.sib.ch).
-----
CC DR EMBL; U30895; AAC14578.1; -.
CC InterPro; IPR000728; -.
CC Pfam; PF00586; AIRS; 1.
KW Purine biosynthesis; Ligase; Chloroplast; Transit peptide.
FT TRANSIT 1 388 CHLOROPLAST (POTENTIAL).
FT CHAIN ? 388 PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-
SQ SEQUENCE 388 AA; 40427 MW; 1SD0EF1127CE9E6 CRC64; LIGASE.
-----
Query Match 9.0%; Score 6; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY 1 GGEFGL 6
Db 93 GGEGL 98
-----
RESULT 24
KICS_HUMAN STANDARD; PRT; 400 AA.
AC P08727;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE KERATIN, TYPE I CYTOSKELETAL 19 (CYTOKERATIN 19) (K19) (CK 19).
GN KRT19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=92325250; PubMed=2469734;
RX Stasiek P.C., Purkis P.E., Leigh I.M., Lane E.B.;
RA "Keratin 19.: predicted amino acid sequence and broad tissue
distribution suggest it evolved from keratinocyte keratins.";
RL J. Invest. Dermatol. 92:707-716(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88096504; PubMed=2447559;
RA Stasiek P.C., Lane E.B.;
RZ "Sequence of cDNA coding for human keratin 19.";
RL Nucleic Acids Res. 15:10058-10058(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=88124986; PubMed=2448790;
RA Eckert R.L.;
RZ "Sequence of the human 40-kDa keratin reveals an unusual structure
with very high sequence identity to the corresponding bovine
keratin";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1114-1118(1988).
RN [4]
RP SEQUENCE FROM N.A.
```

```

RA MEDLINE=20090660; PubMed=10623642;
RX Whitlock N.V., Bady R.A., McGrath J.A.;
RT "Genomic organization and amplification of the human keratin 15 and 19 genes.";
RL Biochem. Res. Commun. 267:462-465(2000).
RN [5]
RP SEQUENCE OF 25-31; 151-158 AND 227-237.
RC TISSUE=Keratinoctyes;
RX MEDLINE=91162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
Vandekekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel protein databases of normal human epidermal keratinoctyes.";
RL Electrophoresis 13:960-969(1992).
CC -I- FUNCTION: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8]. BOTH A BASIC AND AN ACIDIC KERATIN ARE REQUIRED FOR FILAMENT ASSEMBLY.
CC -I- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -I- DOMAIN: THIS KERATIN DIFFERS FROM ALL OTHER IF PROTEINS IN LACKING THE C-TERMINAL TAIL DOMAIN.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/csb-ch) or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00503; CA68556.1; -.
DR EMBL; J03607; AA83604.1; -.
DR EMBL; AF202321; AAF27048.1; -.
DR PIR; A60779; KRH09.
DR SWISS-2DPAGE; P08727; HUMAN.
DR Aarhus/Ghent-2DPAGE; 8216; IEF.
DR MIM; 148020; -.
DR InterPro; IPRO01664; -.
DR InterPro; IPRO02957; -.
DR Pfam; PF00038; filament_1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.
FT FT 1 79 HEAD.
FT DOMAIN 1 ROD.
FT FT 80 387 ROD-LIKE HELICAL TAIL.
FT DOMAIN 388 400 COIL 1A.
FT FT 80 115 LINKER 1.
FT DOMAIN 116 133 COIL 1B.
FT FT 134 225 LINKER 12.
FT DOMAIN 226 248 COIL 2.
FT FT 249 387 STUTTER.
FT SITE 267 267 STUTTER.
FT SITE 327 327 STUTTER.
FT CONFLICT 76 77 LA -> H (IN REF. 2).
FT CONFLICT 342 342 I -> Y (IN REF. 2).
FT CONFLICT 350 350 A -> G (IN REF. 3).
SQ SEQUENCE 400 AA; 44106 MW; 1EBFF9AC82F09CAB5 CRC64;

Query Match          9.0%; Score 6; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 45;
Matches    6; Conservative      0; Mismatches     0; Indels       0; Gaps        0;

OY      3 FGGGCG 8
         |||||
Db      15 FGGGCG 20

RESULT 25
ID      YFJD_ECOLI STANDARD; PRT; 428 AA.
AC      P37908; P76600; P76601; P77009;
```

DT 01-OCT-1994 (Rel. 30, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYDROTHERMAL 46.4 KDA PROTEIN IN FFH-GRE INTERGENIC REGION.  
GN YFJD OR B2612/B2613.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y. ;  
RT "The complete genome sequence of Escherichia coli K-12." ;  
RT Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,  
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K.,  
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,  
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,  
RA Motomura K., Nakamura Y., Nasahimoto H., Nishio Y., Oshima T.,  
RA Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C.,  
RA Yamamoto Y., Yano M. ;  
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 322-428 FROM N.A.  
RC STRAIN=B178;  
RX MEDLINE=88319942; PubMed=3045760;  
RA Lipinska B., King J., Ang D., Georgopoulos C. ;  
RT "Sequence analysis and transcriptional regulation of the Escherichia  
RT coli gypE gene, encoding a heat shock protein." ;  
RL Nucleic Acids Res. 16:7545-7562(1988).  
RN [4]  
RP IDENTIFICATION.  
RC Rudd K.E. ;  
RL Unpublished observations (AUG-1994).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE YPFO053 FAMILY. STRONG, TO  
CC H. INFLUENZAE HI1017.  
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.  
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS  
CC FRAMESHIFTS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement/or-send-an-email-to-licence@isb-sib.ch)  
CC or send an email to [licence@isb-sib.ch](mailto:licence@isb-sib.ch)).  
CC -----  
CC EMBL: AE000347; AAC75661.1; ALT\_FRAME.  
CC EMBL: AE000347; AAC75662.1; ALT\_FRAME.  
CC EMBL: D90888; BA016497.1; ALT\_INIT.  
CC EMBL: X07863; -; NOT\_ANNOTATED\_CDS.  
CC Ecocore: EG12442; YFJD.  
CC InterPro: IPR000644; -  
CC InterPro: IPR002550; -  
CC Pfam: PF00571; CBS; 2.  
CC Pfam: PF01595; DUF21; 1.  
CC Hypothetical protein; Transmembrane.  
CC TRANSMEM 4 24  
CC TRANSMEM 65 85 POTENTIAL.  
CC TRANSMEM 92 112 POTENTIAL.  
CC TRANSMEM 130 150 POTENTIAL.  
CC SEQUENCE. 428 AA; 48044 MW; 38456865EDBCB151 CRC64;

Query Match	9.0%;	Score 6;	DB 1;	length 428;
Best Local Similarity	100.0%;	Pred. No. 47;		
Matches	6;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Q7	54	RNKKKV	59	
Db	297	RNKKKV	302	

Search completed: July 12, 2001, 06:23:08  
Job time: 334 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 06:16:45 ; Search time 14.91 Seconds

(without alignments)  
342.300 Million cell updates/sec

Title: US-09-506-978-1

Sequence: 1 GGRGGGCGGCGKCPSENEIFSR.....CRILGYLRNKKKVCVPSKCG 67

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 50 summaries

Database :

1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8	11.9	139	2	B33910	sal homeotic prote
2	8	11.9	151	2	T08002	glycine-rich prote
3	8	11.9	543	2	F70726	hypothetical glyci
4	8	11.9	1329	2	E70917	hypothetical glyci
5	7	10.4	100	2	T45643	hypothetical prote
6	7	10.4	284	2	T23158	hypothetical prote
7	7	10.4	300	1	S16815	SNP1 protein - yea
8	7	10.4	314	2	T30523	dihydroorotate deh
9	7	10.4	360	1	S35981	L-iditol 2-dehydro
10	7	10.4	378	2	S72599	hypothetical prote
11	7	10.4	382	2	D70587	hypothetical prote
12	7	10.4	438	2	B75563	GRP-binding protei
13	7	10.4	450	2	S73419	signal recognition
14	7	10.4	450	2	F71351	probable preprotel
15	7	10.4	463	2	F70760	probable dlnx prot
16	7	10.4	465	2	B70213	hypothetical prote
17	7	10.4	622	2	I37984	keratin 9, type I,
18	7	10.4	652	2	T03504	probable DNA topol
19	7	10.4	682	2	T28899	hypothetical prote
20	7	10.4	756	2	F71340	probable tpr prote
21	7	10.4	758	2	F71301	probable tpr prote
22	7	10.4	762	2	C71340	probable tpr prote
23	7	10.4	856	2	G71133	probable alpha-man
24	7	10.4	944	2	T28734	hypothetical prote
25	7	10.4	1804	2	H96597	hypothetical prote
26	6	9.0	35	1	NTSRPM	neurotoxin p2 - sc
27	6	9.0	78	2	B70909	hypothetical prote
28	6	9.0	79	2	T09851	hypothetical prote
29	6	9.0	89	2	B84145	hypothetical prote

30	6	9.0	140	2	S72630	ribosomal protein
31	6	9.0	148	2	H72781	hypothetical prote
32	6	9.0	158	2	B54762	phospholipase A2 (
33	6	9.0	188	2	B82471	hypothetical prote
34	6	9.0	203	2	D64885	superinfectio exc
35	6	9.0	203	2	H85747	hypothetical prote
36	6	9.0	213	2	T22900	hypothetical prote
37	6	9.0	227	1	H65039	hypothetical prote
38	6	9.0	227	1	G85907	probable transport
39	6	9.0	249	2	T09407	stationary-phase s
40	6	9.0	256	2	G69997	hypothetical prote
41	6	9.0	261	2	T37948	probable ui small
42	6	9.0	268	2	B61615	fibroin heavy chai
43	6	9.0	274	2	A61030	nitrogenase (nc 1,
44	6	9.0	277	2	C82128	4-amino-4-deoxycho
45	6	9.0	277	2	F75548	conserved hypochet
46	6	9.0	284	2	D84057	hypothetical prote
47	6	9.0	288	2	A83443	probable transcrip
48	6	9.0	292	2	C84865	hypothetical prote
49	6	9.0	293	1	R3R52	ribosomal protein
50	6	9.0	300	2	T49933	inorganic pyrophos

#### ALIGNMENTS

RESULT 1  
B33910  
sal homeotic protein - fruit fly (*Drosophila simulans*)  
C:Species: *Drosophila simulans*  
C>Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 20-Aug-1999  
C:Accession: B33910  
R:Reuter, D.; Schuh, R.; Jaeckle, H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 5483-5486, 1989  
A>Title: The homeotic gene spalt (sal) evolved during *Drosophila* speciation.  
A:Reference number: A33910; MOID:89315821  
A:Accession: B33910  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-139 <RED>  
A:Cross-references: GB:M21227; NID:q158375; PIDN:AAA28977.1; PID:q158376  
C:Genetics:  
A:Gene: FlyBase: Dsim/sala  
A:Cross-references: FlyBase:FBgn0012892  
C:Superfamily: homeotic protein sal

Query Match 11.9%; Score 8; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGRGGGCG 8  
DB 32 GGRGGGCG 39

#### RESULT 2

T08002  
glycine-rich protein - common safinoin  
C:Species: *Onobrychis viciifolia* (common safinoin)  
C>Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999  
C:Accession: T08002  
R:Joseph, R.G.  
submitted to: The EMBL Data Library, October 1997  
A:Description: A cDNA coding for a glycine-rich protein from *Onobrychis viciifolia*.  
A:Reference number: Z16281  
A:Accession: T08002  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-151 <JOS>  
A:Cross-references: EMBL:AF027686; NID:q2565428; PIDN:AA82000.1; PID:q2565429

Query Match 11.9%; Score 8; DB 2; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 0.39;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 |||||  
 DB 76 GGFGLG 83

## RESULT 3

hypothetical glycine-rich protein RV2591 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: F70726  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987

A:Accession: F70726

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-543 <COL>

A:Cross-references: GB:277724; GB:AL123456; NID:93261620; PIDN:CAB01283.1; PID:91478221

A:Experimental source: strain H37RV  
 C:Genetics:

A:Gene: RV2591  
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 11.9%; Score 8; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 |||||  
 DB 380 GGFGLG 387

## RESULT 4

hypothetical glycine-rich protein RV1450c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: E70917

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987

A:Accession: E70917

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-1329 <COL>

A:Cross-references: GB:295844; GB:AL123456; NID:93250713; PIDN:CAB09271.1; PID:92131046

A:Experimental source: strain H37RV  
 C:Genetics:

A:Gene: RV1450c  
 C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 11.9%; Score 8; DB 2; Length 1329;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 8  
 |||||  
 DB 1318 GGFGLG 1325

## RESULT 5

hypothetical protein F13112.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000

C:Accession: T45643  
 R:Choisme, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Arliguene, F.;

submitted to the Protein Sequence Database, November 1999  
 A:Reference number: 223010

A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-100 <CHO>

A:Cross-references: EMBL:AL133292  
 A:Experimental source: cultivar Columbia; BAC clone F13112

C:Genetics:

A:Map position: 3  
 A:Note: F13112.120

C:Superfamily: Arabidopsis thaliana hypothetical protein F13112.120

Query Match 10.4%; Score 7; DB 2; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 7  
 |||||  
 DB 89 GGFGLG 95

## RESULT 6

hypothetical protein K01A6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T23158  
 R:Collage, A.

submitted to the EMBL Data Library, January 1996  
 A:Reference number: Z19701

A:Accession: T23158

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-284 <WIL>

A:Cross-references: EMBL:268750; PIDN:CA92963.1; GSPDB:GN00022; CESP:K01A6.4

A:Experimental source: clone K01A6  
 C:Genetics:

A:Gene: CESP:K01A6.4

A:Map position: 4  
 A:Insertions: 89/3; 104/3; 126/3

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 10.4%; Score 7; DB 2; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 7.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 7  
 |||||  
 DB 259 GGFGLG 265

## RESULT 7

SNP1 protein - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: S16815; S46418  
 R:Smith, V.; Barrell, B.G.

EMBO J. 10, 2627-2634, 1991  
 A:Title: Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation o

A:Reference number: S16815; MUID:91330888  
A:Accession: S16815  
A:Molecule type: DNA  
A:Residues: 1-300 <SM1>  
A:Cross-references: EMBL:X59986; NID:94503; PIDN:CAA42602.1; PID:94504  
A:Experimental source: strain AB972  
R:Smith, V.  
submitted to the EMBL Data Library, September 1994  
A:Reference number: S48407  
A:Accession: S48418  
A:Molecule type: DNA  
A:Residues: 1-300 <SM2>  
A:Cross-references: GB:247047; EMBL:238060; NID:9603997; PID:9763285; GSPDB:GN00009; MIF  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:SNP1; MIPS:YIL061C  
A:Cross-references: SGD:S0001323; MIPS:YIL061C  
A:Map position: 9L  
C:Superfamily: yeast SNP1 protein; ribonucleoprotein repeat homology  
C:Keywords: nucleus  
F:108-178/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 10.4%; Score 7; DB 1; Length 300;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGRC 10  
|||||  
DB 201 GGLGRC 207

RESULT 8  
T30523  
dihydroorotate dehydrogenase (EC 1.3.99.11) - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T30523  
R:Go, G.; Nara, T.; Nakajima-Shimada, J.; Aoki, T.  
J. Mol. Biol. 285, 149-161, 1999  
A:Title: Novel organization and sequences of five genes encoding all six enzymes for de  
A:Reference number: Z20841; MUID:99096912  
A:Accession: T30523  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-314 <GA0>  
A:Cross-references: EMBL:AB017765; NID:95509902; PIDN:BAA74526.1; PID:94210458  
C:Genetics:  
A:Note: PYR4  
C:Superfamily: dihydroorotate oxidase  
C:Keywords: oxidoreductase

Query Match 10.4%; Score 7; DB 2; Length 314;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLG 8  
|||||  
DB 218 GFGGLG 224

RESULT 9  
S35981  
L-iditol 2-dehydrogenase (EC 1.1.1.14) - fission yeast (Schizosaccharomyces pombe)  
N:Alternate names: sorbitol dehydrogenase  
C:Species: Schizosaccharomyces pombe  
C:Date: 13-Jan-1995 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
C:Accession: T39670; S38345; S35981  
R:Lyne, M.; Rajandream, M.A.; Bartell, B.G.; Beck, A.; Borzym, K.; Klages, S.; Langer, T.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z21847  
A:Accession: T39670

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-360 <LYN>  
A:Cross-references: EMBL:AL033388; PIDN:CAA21910.1; GSPDB:GN00067; SPDB:SPBC1773.05C  
A:Experimental source: strain 972h-; cosmid cl773  
R:Wagner, P.; Grimaldi, M.; Jenkins, J.R.  
Eur. J. Biochem. 217, 731-736, 1993  
A:Title: Putative dehydrogenase tms1 suppresses growth arrest induced by a p53 tumour  
A:Reference number: S38345; MUID:94039112  
A:Accession: S38345  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 14-360 <MA2>  
A:Cross-references: EMBL:X74422; NID:9396478; PIDN:CAA52443.1; PID:9396479  
A:Note: submitted to the EMBL Data Library, July 1993  
C:Genetics:  
A:Gene: SPDB:SPBC1773.05c; tms1  
A:Map position: 2  
C:Complex: homotetramer  
C:Function:  
A:Description: catalyzes the oxidation by NAD+ of sorbitol to fructose  
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
C:Keywords: homotetramer; metalloprotein; NAD; oxidoreductase; zinc  
F:27-346/Domain: long-chain alcohol dehydrogenase homology <LADH>  
F:172-201/Region: beta-alpha-beta NAD nucleotide-binding fold  
F:42,67,153/Binding site: zinc, catalytic (Cys, His, Glu) #status predicted  
F:97,100,103,111/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 10.4%; Score 7; DB 1; Length 360;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 PGVCRL 50  
|||||  
DB 93 PGVCRL 99

RESULT 10  
S72599  
hypothetical protein B1937\_F2\_56 - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Aug-1999  
C:Accession: S72599  
R:Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, November 1993  
A:Reference number: S72580  
A:Accession: S72599  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <SM1>  
A:Cross-references: EMBL:U00016; NID:9466961; PIDN:AA17167.1; PID:9466981  
C:Genetics:  
A:Start codon: GTC  
A:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F:4-68/Domain: dnaJ amino-terminal homology <DND>

Query Match 10.4%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GFGGLG 7  
|||||  
DB 82 GFGGLG 88

RESULT 11  
D70587  
probable dnaJ protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70587

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: D70587  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1382 <COL>  
 A: Cross-references: GB:295208; GB:AL123456; NID:93261747; PIDN:CA08479.1; PID:92078026  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: dnaJ2  
 C: Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
 F: 4-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 10.4%; Score 7; DB 2; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGFGLG 7  
 |||||  
 DB 82 GGFGLG 88

RESULT 12  
 B75563  
 GTP-binding protein Obg - *Deinococcus radiodurans* (strain R1)  
 C: Species: *Deinococcus radiodurans*  
 C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C: Accession: B75563  
 R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.; Lam, P.; McDonald, L.; Uitterback, T.; Zaleski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A: Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A: Reference number: A75250; MUID: 20036896  
 A: Accession: B75563  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1438 <WHI>  
 A: Cross-references: GB:AE001871; GB:AE000513; NID:96457740; PIDN:AAF09676.1; PID:9645774  
 A: Experimental source: strain R1  
 C: Genetics:  
 A: Gene: DR0084  
 A: Map position: 1  
 C: Superfamily: GTP-binding protein obg; translation elongation factor Tu homology

Query Match 10.4%; Score 7; DB 2; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGGRG 10  
 |||||  
 DB 120 GGLGGRG 126

RESULT 13  
 S73419  
 signal recognition particle protein fth - *Mycoplasma pneumoniae* (strain ATCC 29342)  
 N: Alternate names: hypothetical protein D09\_07450  
 C: Species: *Mycoplasma pneumoniae*  
 A: Variety: ATCC 29342  
 C: Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C: Accession: S73419  
 R: Himmelfeich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A: Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*  
 A: Reference number: S73327; MUID: 97105885

A: Accession: S73419  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1450 <HIM>  
 A: Cross-references: EMBL:AE000011; GB:U00089; NID:91673740; PIDN:AA095741.1; PID:9167  
 A: Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C: Genetics:  
 A: Gene: fth  
 A: Genetic code: SGC3  
 C: Superfamily: signal recognition particle 54k protein

Query Match 10.4%; Score 7; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLG 8  
 |||||  
 DB 441 GFGGLG 447

RESULT 14  
 F71351  
 probable preprotein translocase subunit (secY) - *Syphilis spirochete*  
 C: Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)  
 C: Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 C: Accession: F71351  
 R: Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
 rson, J.; Khailak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uitterback, T.; M  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A: Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.  
 A: Reference number: A71250; MUID: 98332770  
 A: Accession: F71351  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1450 <COL>  
 A: Cross-references: GB:AE001203; GB:AE000520; NID:93322476; PIDN:AC65198.1; PID:9332  
 A: Experimental source: strain Nichols  
 C: Genetics:  
 A: Gene: pp0208  
 C: Superfamily: preprotein translocase secY

Query Match 10.4%; Score 7; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GLGGRG 11  
 |||||  
 DB 427 GLGGRG 433

RESULT 15  
 F70760  
 probable dlnx protein - *Mycobacterium tuberculosis* (strain H37RV)  
 C: Species: *Mycobacterium tuberculosis*  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C: Accession: F70760  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: F70760  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1463 <COL>  
 A: Cross-references: GB:274020; GB:AL123456; NID:93261584; PIDN:CAA98325.1; PID:924855  
 A: Experimental source: strain H37RV  
 C: Genetics:



A:Gene: dlnx

Query Match 10.4%; Score 7; DB 2; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGGRG 10  
 |||||  
 DB 29 GGLGGRG 35

RESULT 16

B70213

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/lp54

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C&gt;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 28-Jul-2000

C:Accession: B70213

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A&gt;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943

A:Accession: B70213

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-465 &lt;KLE&gt;

A:Cross-references: GB:AE000790; NID:92630224; PIDN:AAC66282.1; PID:g2690281; TIGR:BBA50

A:Experimental source: strain B31

C:Genetics:

A:Genome: plasmid

C:Superfamily: Borrelia burgdorferi hypothetical protein BBA50

Query Match

Best Local Similarity 10.4%; Score 7; DB 2; Length 465;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLGG 8

|||||

DB 44 GFGGLGG 50

RESULT 17

I37984

Keratin 9, type I, cytoskeletal - human

N:Alternate names: cytokeratin 9; scatter protein 60K chain, placental

C:Species: Homo sapiens (man)

C&gt;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 29-Oct-1999

C:Accession: I37984; S40307; S77921; S41161; B35494; I37943

R:Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Schroe

Nature Genet. 6, 174-179, 1994

A&gt;Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).

A:Reference number: I37984; MUID:94214498

A:Accession: I37984

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-622 &lt;RES&gt;

A:Cross-references: EMBL:X75015; NID:9453154; PIDN:CAA52924.1; PID:g453155

R:Langbein, L.; Held, H.W.; Moll, I.; Franke, W.W.

Differentiation 55, 57-71, 1993

A&gt;Title: Molecular characterization of the body site-specific human epidermal cytokerati

A:Reference number: I37943; MUID:94131202

A:Accession: S40307

A:Molecule type: mRNA

A:Residues: 1-622 &lt;LAN&gt;

A:Cross-references: EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PID:g435476

A:Accession: S77921

A:Molecule type: protein

A:Residues: 29-53; 62-104; 167-188; 199-233; 241-249; 295-362; 374-430; 450-480; 579-604 &lt;LAF&gt;

R:Langbein, L.

submitted to the EMBL Data Library, December 1993

A:Reference number: S41161

A:Accession: S41161

A:Molecule type: mRNA

A:Residues: 1-11, 'SR', 13-622 &lt;LAN&gt;

A:Cross-references: EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PID:g435476

R:Rosen, E.M.; Meromsky, L.; Romero, R.; Settle, E.; Goldberg, I.

Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990

A&gt;Title: Human placenta contains an epithelial scatter protein.

A:Reference number: A35494; MUID:90267446

A:Accession: B35494

A:Molecule type: protein

A:Residues: 'X', 450-465 &lt;ROS&gt;

C:Genetics:

A:Gene: GDB:KRT9; EPPK

A:Cross-references: GDB:303970; OMIM:144200

A:Map position: 17q12-17q21

A:Insertions: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2

A&gt;Note: defects in this gene may cause epidermolytic palmoplantar keratoderma

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

F:1-153/Domain: head #status predicted &lt;HEA&gt;

F:154-459/Domain: helical rod #status predicted &lt;ROD&gt;

F:460-622/Domain: tail #status predicted &lt;TAI&gt;

Query Match

Best Local Similarity 10.4%; Score 7; DB 2; Length 622;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGGLGG 8

|||||

DB 129 GFGGLGG 135

RESULT 18

T03504

probable DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain B - Rhodobacter caps

N:Alternate names: DNA gyrase chain b

C:Species: Rhodobacter capsulatus

C&gt;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C:Accession: T03504

R:VLcek, C.; Paces, V.; Matsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.

Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A&gt;Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1

A:Reference number: Z14955; MUID:97404404

A:Accession: T03504

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-652 &lt;VLC&gt;

A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16157.1; PID:g3128305

C:Genetics:

A:Map position: 1

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

C:Keywords: isomerase

Query Match 10.4%; Score 7; DB 2; Length 652;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GLGGRGK 11

|||||

DB 573 GLGGRGK 579

RESULT 19

T28899

hypothetical protein T17H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T28899

R:Favellio, A.

submitted to the EMBL Data Library, January 1996  
A:Description: The sequence of *C. elegans* cosmid T17H7.

A:Reference number: Z20540

A:Accession: T28899

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-682 <FAV>

A:Cross-references: EMBL:U42841; PIDN:AAC48170.1; GSPDB:GN00021; CESP:T17H7.1

A:Experimental source: strain Bristol N2; clone T17H7

C:Genetics:

A:Gene: CESP:T17H7.1

A:Map position: 3

A:Introns: 85/3; 147/3

Query Match 10.4%; Score 7; DB 2; Length 682;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGLGRC 10  
|||||||  
Db 587 GGLGRC 593

#### RESULT 20

Probable tpr protein G (tprG) - *Syphilis spirochete*

C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: F71340

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

son, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McD

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.

A:Reference number: A71250; MUID:98332770

A:Accession: F71340

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-756 <COL>

A:Cross-references: GB:AE001211; GB:AE000520; NID:g3322582; PIDN:AAC65301.1; PID:g332256

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0317

Query Match 10.4%; Score 7; DB 2; Length 756;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGLGG 8  
|||||||  
Db 169 GFGLGG 175

#### RESULT 21

Probable tpr protein J (tprJ) - *Syphilis spirochete*

C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: F71301

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

son, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McD

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.

A:Reference number: A71250; MUID:98332770

A:Accession: F71301

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-758 <COL>

A:Cross-references: GB:AE001237; GB:AE000520; NID:g3322916; PIDN:AAC65595.1; PID:g332291

A:Experimental source: strain Nichols

C:Genetics:  
A:Gene: TP0621

Query Match 10.4%; Score 7; DB 2; Length 758;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGLGG 8  
|||||||  
Db 169 GFGLGG 175

#### RESULT 22

Probable tpr protein E (tprE) - *Syphilis spirochete*

C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: C71340

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

winson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; M

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.

A:Reference number: A71250; MUID:98332770

A:Accession: C71340

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-762 <COL>

A:Cross-references: GB:AE001211; GB:AE000520; NID:g3322582; PIDN:AAC65300.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0313

Query Match 10.4%; Score 7; DB 2; Length 762;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GFGLGG 8  
|||||||  
Db 169 GFGLGG 175

#### RESULT 23

Probable alpha-mannosidase - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999

C:Accession: G71133

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Ogi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: G71133

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-856 <RAW>

A:Cross-references: GB:AP000003; NID:g3336130; PIDN:BA29929.1; PID:d1030872; PID:g32

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by Genba

C:Genetics:

A:Gene: PH0835

Query Match 10.4%; Score 7; DB 2; Length 856;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 LRNRKV 59  
|||||||  
Db 713 LRNRKV 719

## RESULT 24

T28734

hypothetical protein F26G5.9 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T28734

R:Sammons, L.; Wohlmann, P.; Beck, C.

A:Submitted to the EMBL Data Library, September 1997

A:Description: The sequence of *C. elegans* cosmid F26G5.

A:Reference number: Z20516

A:Accession: T28734

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-944 &lt;SAM&gt;

A:Cross-references: EMBL:AF022974; PIDN:AAC48041.1; GSPDB:GN00023; CESP:F26G5.9

A:Experimental source: strain Bristol N2; clone F26G5

C:Genetics:

A:Gene: CESP:F26G5.9

A:Map position: 5

A:introns: 12/3; 48/3; 258/3; 327/3; 643/1; 916/2

## Query Match

10.4%; Score 7; DB 2; Length 944;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGGIGG 8

|||||

Db 76 GFGGIGG 82

## RESULT 25

H96597

hypothetical protein TSA14.5 [imported] - *Arabidopsis thaliana*C:Species: *Arabidopsis thaliana* (mouse-ear cross)

C&gt;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: H96597

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: AB6141; MUID:21016719

A:Accession: H96597

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1804 &lt;STO&gt;

A:Cross-references: GB:AE005173; NID:g4204261; PIDN:AAD10642.1; GSPDB:GN00141

C:Genetics:

A:Gene: TSA14.5

A:Map position: 1

## Query Match

10.4%; Score 7; DB 2; Length 1804;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFGGIGG 7

|||||

Db 1733 GFGGIGG 1739

Search completed: July 12, 2001, 06:17:29  
 Job time: 44 sec

